(1)

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 30, 2002, 22:13:59 ; Search time 57.04 Seconds
(without alignments)
564.899 Million cell updates/sec

Title: US-09-578-063-47
Perfect score: 2247
Sequence: 1 MLETLSRQWIVSHRMEMWLL.....IHLMQQEETNLSQGRCEAVL 423

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1290	57.4	399	2	S41408	lysosomal acid lip
2	1290	57.4	399	2	G01416	lysosomal acid lip
3	1164.5	51.8	395	1	LIRTT	triacylglycerol li
4	1139	50.7	397	1	JC4017	triacylglycerol li
5	1119	49.8	398	2	S07145	triacylglycerol li
6	752.5	33.5	403	2	T33198	hypothetical prote
7	751	33.4	411	2	T22290	hypothetical prote
8	730	32.5	405	2	T22675	hypothetical prote
9	655.5	29.2	426	2	T20480	hypothetical prote
10	574.5	25.6	559	2	JT0949	egg-specific prote
11	505	22.5	443	2	T39540	triglyceride lipas
12	478.5	21.3	1585	2	T31611	hypothetical prote
13	475	21.1	467	2	T41053	triglyceride lipas
14	445	19.8	344	2	E84526	probable lysosomal
15	422.5	18.8	548	2	S37969	probable triacylgl
16	412.5	18.4	460	2	T39443	probable triglycer
17	411.5	18.3	413	2	T43170	probable triacylgl
18	294	13.1	431	2	D86318	protein F15H18.6 [
19	285	12.7	538	2	S64842	probable membrane
20	271.5	12.1	573	2	S64754	probable membrane
21	234.5	10.4	509	2	G96766	protein lipase F2P
22	149	6.6	62	2	S59904	lipase, pregastric
23	143	6.4	336	2	F83425	probable esterase/
24	126	5.6	987	2	A64474	hypothetical prote
25	113	5.0	412	2	B72391	conserved hypothet
26	112.5	5.0	40	2	S19539	triacylglycerol li
27	105	4.7	430	2	T23001	hypothetical prote
28	103.5	4.6	798	2	F96714	probable protease
29	101.5	4.5	929	2	JH0262	QutR protein - Eme
30	101	4.5	436	2	JC5021	platelet-activatin
31	99.5	4.4	286	2	D75217	probable 2-acetyl-
32	99.5	4.4	1280	2	B34087	hypothetical prote
33	99	4.4	460	2	G96764	unknown protein F2
34	97.5	4.3	322	2	B70957	probable ephA prot
35	97.5	4.3	456	2	E69391	hypothetical prote
36	96.5	4.3	287	1	E64053	hypothetical prote
37	96.5	4.3	1275	2	I38588	reverse transcript
38	96	4.3	809	2	S67153	probable membrane
39	96	4.3	934	2	B82364	DNA polymerase I V
40	95.5	4.3	1259	4	GNHUL1	retrovirus-related
41	95	4.2	973	2	T41201	isoleucyl-trna syn
42	94.5	4.2	561	2	G83913	hypothetical prote
43	93	4.1	1275	2	B28096	line-1 protein ORF
44	93	4.1	2311	1	TVCHSR	kinase-related pro
45	92.5	4.1	1275	2	S65824	reverse transcript

ALIGNMENTS

RESULT 1
S41408
lysosomal acid lipase (EC 3.1.1.-) / sterol esterase (EC 3.1.1.13) precursor - human
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 18-Jun-1999
C;Accession: S41408; A39315; S47187
R;Ameis, D.; Merkel, M.; Eckerskorn, C.; Greten, H.
Eur. J. Biochem. 219, 905-914, 1994
A;Title: Purification, characterization and molecular cloning of human hepatic lysoso
A;Reference number: S41408; MUID:94155897
A;Accession: S41408
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-399 <AME>
A;Cross-references: EMBL:X76488; NID:g434305; PIDN:CAA54026.1; PID:g434306
R;Anderson, R.A.; Sando, G.N.
J. Biol. Chem. 266, 22479-22484, 1991
A;Title: Cloning and expression of cDNA encoding human lysosomal acid lipase/choleste
A;Reference number: A39315; MUID:92042192
A;Accession: A39315
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-15,'P',17-399 <AND>
A;Cross-references: GB:M74775; NID:g187151; PIDN:AAA59519.1; PID:g187152
R;Du, H.; Gregory, G.A.
submitted to the EMBL Data Library, April 1994
A;Description: Structural conservation of putative functional motifs between mouse an
A;Reference number: S47187
A;Accession: S47187
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-22,'R',24-399 <DUH>
A;Cross-references: EMBL:Z31690; NID:g506430; PIDN:CAA83495.1; PID:g506431
C;Genetics:
A;Gene: GDB:LIPA
A;Cross-references: GDB:120153; OMIM:278000
A;Map position: 10q24-10q25
C;Superfamily: triacylglycerol lipase, lingual
C;Keywords: carboxylic ester hydrolase; glycoprotein

Query Match 57.4%; Score 1290; DB 2; Length 399;
Best Local Similarity 61.0%; Pred. No. 6.1e-106;
Matches 242; Conservative 59; Mismatches 94; Indels 2; Gaps 1;

```
Qy      15 MEMWLLILVAYMFQRNVNSVHMPTK--AVDPEAFMNISEIIQHQGYPCEEYEVATEDGYI 72
           |:|  | ||  :    ::|    |  ||||| ||:|||| : |:| ||| | ||||||
Db       1 MKMRFLGLVVCLVLWTLHSEGSGGKLTAVDPETNMNVSEIISYWGFPSEEYLVETEDGYI 60

Qy      73 LSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNS 132
           | :|||| |       | :||| |||||:  :|||::|| |:|||||||||||||||||
Db      61 LCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANSSLGFILADAGFDVWMGNS 120
```

```
Qy    133 RGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMG 192
          ||| |||||||||: |||||||||||||::|||| ||||| |||||::||||:||||||:|
Db    121 RGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFILNKTGQEQVYYVGHSQGTTIG 180

Qy    193 FIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFL 252
          ||||| :||||::|||:||| |:|:|    ||  |   ||| :|| ||| |||| |: ||
Db    181 FIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDLFGDKEFLPQSAFL 240

Qy    253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
          : |  ::|  ||| ::| |:  || |||  |:|||  ||  |: |||||||:|||||||
Db    241 KWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSRVDVYTTHSPAGTSVQNMLHWSQAV 300

Qy    313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
          : :||||||  ||    ||  |  | |:|| ||||:|:|| |||::  || :||:::
Db    301 KFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDMLVPTAVWSGGHDWLADVYDVNILLTQI 360

Qy    373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQ 409
          |||::|::|||| |:|||||||| |:||:||:||::
Db    361 TNLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRK 397


RESULT   2
G01416
lysosomal acid lipase - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 18-Jun-1999
C;Accession: G01416
R;Du, H.
submitted to the EMBL Data Library, April 1994
A;Reference number: G06919
A;Accession: G01416
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-399 <DUX>
A;Cross-references: EMBL:U08464; NID:g505052; PIDN:AAB60328.1; PID:g505053
C;Superfamily: triacylglycerol lipase, lingual
C;Keywords: glycoprotein


  Query Match             57.4%;  Score 1290;  DB 2;  Length 399;
  Best Local Similarity   61.0%;  Pred. No. 6.1e-106;
  Matches  242;  Conservative   59;  Mismatches   94;  Indels    2;  Gaps    1;

Qy     15 MEMWLLILVAYMFQRNVNSVHMPTK--AVDPEAFMNISEIIQHQGYPCEEYEVATEDGYI 72
          |:|  | ||  :    ::|     |  ||||| ||:|||| : |:| ||| | ||||||
Db      1 MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYWGFPSEEYLVETEDGYI 60

Qy     73 LSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNS 132
          | :|||| |       | :||| |||||:  :|||::|| |:|||||||||||||||||||
Db     61 LCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANSSLGFILADAGFDVWMGNS 120

Qy    133 RGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMG 192
          ||| |||||||||: |||||||||||||::|||| ||||| |||||::||||:||||||:|
Db    121 RGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFILNKTGQEQVYYVGHSQGTTIG 180

Qy    193 FIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFL 252
          ||||| :||||::|||:||| |:|:|    ||  |   ||| :|| ||| |||| |: ||
Db    181 FIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDLFGDKEFLPQSAFL 240

Qy    253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
          : |  ::|  ||| ::| |:  || |||  |:|||  ||  |: |||||||:|||||||
Db    241 KWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSRVDVYTTHSPAGTSVQNMLHWSQAV 300

Qy    313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
          : :||||||  ||    ||  |  | |:|| ||||:|:|| |||::  || :||:::
Db    301 KFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDMLVPTAVWSGGHDWLADVYDVNILLTQI 360

Qy    373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQ 409
          |||::|::|||| |:|||||||| |:||:||:||::
Db    361 TNLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRK 397


RESULT   3
LIRTT
triacylglycerol lipase (EC 3.1.1.3) precursor, lingual - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 11-Jun-1999
C;Accession: A23045
R;Docherty, A.J.P.; Bodmer, M.W.; Angal, S.; Verger, R.; Riviere, C.; Lowe, P.A.; L
Nucleic Acids Res. 13, 1891-1903, 1985
A;Title: Molecular cloning and nucleotide sequence of rat lingual lipase cDNA.
A;Reference number: A23045; MUID:85215587
A;Accession: A23045
A;Molecule type: mRNA
A;Residues: 1-395 <DOC>
A;Cross-references: GB:X02309; NID:g56595; PIDN:CAA26179.1; PID:g56596
A;Experimental source: strain Sprague-Dawley
A;Note: the partial sequence of the mature protein from a different, unspecified str
C;Comment: This acid-stable lipase is secreted by the serous (von Ebner's) glands at
C;Superfamily: triacylglycerol lipase, lingual
C;Keywords: carboxylic ester hydrolase; glycoprotein; lipid digestion; saliva; serou
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-395/Product: triacylglycerol lipase, lingual #status predicted <MPT>
F;33,68,98,184,270/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             51.8%;  Score 1164.5;  DB 1;  Length 395;
  Best Local Similarity   54.9%;  Pred. No. 7.1e-95;
  Matches  217;  Conservative   69;  Mismatches  108;  Indels    1;  Gaps    1;

Qy     17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
          ||||:: : :      : :       :||| ||||::| : ||||:|||| ||||||| |
Db      1 MWLLLITSVISTFGGAHGLFGKLGPGNPEANMNISQMITYWGYPCQEYEVVTEDGYILGV 60

Qy     76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
          ||| |    :  | |||| |||||:  |:|||:|||||||| |:|||||:|||:||||||
Db     61 YRIPHGKNNSENIGKRPVVYLQHGLIASATNWIANLPNNSLAFMLADAGYDVWLGNSRGN 120

Qy    136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
          ||||:   | |  ||||||:||||::|||| ||||:|||||||||:|||:||||||:||||
Db    121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATINFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy    196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQL 255
          ||| | ||:||| ::|||||:||||: :||  |   :|  : | :|||| ||  | |   |
Db    181 FSTNPTLAKKIKTFYALAPVATVKYTQSPLKKISFIPTFLFKLMFGKKMFLPHTYFDDFL 240

Qy    256 VIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSG 315
          :| : :|| :||| : :  ||:  |:|:||  ||  |  ||||||: |||:| | ||
Db    241 GTEVCSREVLDLLCSNTLFIFCGFDKKNLNVSRFDVYLGHNPAGTSVQDFLHWAQLVRSG 300

Qy    316 ELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNL 375
          : :||:||| ::|:   || ||  | |  |||| |:| || | |::|:|| ||| :::||
Db    301 KFQAFNWGSPSQNMLHYNQKTPPEYDVSAMTVPVAVWNGGNDILADPQDVAMLLPKLSNL 360

Qy    376 IYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
          ::|| |  : |:|||| :|||  :|||:| :| ::
Db    361 LFHKEILAYNHLDFIWAMDAPQEVYNEMISMMAED 395


RESULT   4
JC4017
triacylglycerol lipase (EC 3.1.1.3) PGE precursor - bovine
N;Alternate names: pregastric esterase
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: JC4017; S64678
R;Timmermans, M.Y.J.; Teuchy, H.; Kupers, L.P.M.
Gene 147, 259-262, 1994
A;Title: The cDNA sequence encoding bovine pregastric esterase.
A;Reference number: JC4017; MUID:95011625
A;Accession: JC4017
A;Molecule type: mRNA
A;Residues: 1-397 <TIM>
A;Cross-references: GB:L26319; NID:g600756; PIDN:AAA57037.1; PID:g600757
A;Experimental source: tongue
```

```
alignment_block:
 US-09-578-063-47 x HSLAL   ..

 Align seg 1/1  to: HSLAL  from: 1  to: 2626

       2 LeuGluThrLeuSerArgGlnTrpIleValSerHisArgMetGluMetTr 18
         |||||||||   :::||||||          |||   ||||||:::|||
     113 CTCGAGACAGCGGCCCGGCAG......GACAGCTCCAGAATGAAAATGCG 156

      18 pLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAsnSerValH 35
          :::|||   ||||||:::  :::          ::::::::::|||
     157 GTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTCTGAGG 206

      35 isMetProThrLys......AlaValAspProGluAlaPheMetAsnIle 49
                    |||      ||||||||||||||:::   ||||||:::
     207 GGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTG 256

      50 SerGluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAl 66
         ||||||||||||:::::   |||:::|||   |||||||||   |||
     257 AGTGAAATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGA 306

      66 aThrGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuV 83
          ||||||||||||||||||||   :::||||||||||||:::|||
     307 GACAGAAGATGGATATATTCTGTGCCTTAACCGAATTCCTCATGGGAGGA 356

      83 alGlnProLysLysThrGlySerArgProValValLeuLeuGlnHisGly 99
           :::   :::      |||   :::||||||||||:::||||||||||||
     357 AGAACCATTCTGACAAAGGTCCCAAACCAGTTGTCTTCCTGCAACATGGC 406

     100 LeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLe 116
         |||::::::   :::||||||||||:::::::||||||   |||:::|||||
     407 TTGCTGGCAGATTCTAGTAACTGGGTCACAAACCTTGCCAACAGCAGCCT 456

     116 uGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSerA 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     457 GGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAACAGCA 506

     133 rgGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAsp 149
         ||||||||:::||||||||||||||||||||||||||||:::::::||||||
     507 GAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGAT 556

     150 GluPheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeuProAl 166
         ||||||||||||||||||||||||||||||||||:::::::||||||||||
     557 GAATTCTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGC 606

     166 aValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrV 183
         |   ||||||||||||||:::||||||||||||||:::::::|||||||
     607 TTCCATTAACTTCATTCTGAATAAAACTGGCCAAGAACAAGTGTATTATG 656

     183 alGlyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMet 199
         |||||:::||||||||||||||:::||||||||||||||||||   :::
     657 TGGGTCATTCTCAAGGCACCACTATAGGTTTTATAGCATTTTCACAGATC 706

     200 ProGluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProIleAl 216
         ||||||||||:::::::|||||||||:::||||||||||:::|||:::||
     707 CCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGCCCTGGGTCCTGTGGC 756

     216 aThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeuProA 233
         |:::|||      :::   ||||||   :::|||:::      |||||||
     757 TTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATTACCAG 806

     233 spMetMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThr 249
         ||   :::||||||   |||||||||   ||||||||||||   |||:::
     807 ATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGT 856

     250 ArgPheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAs 266
            ||||||:::   |||      :::::::|||   :::||||||||||
     857 GCGTTTTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAA 906

     266 pGlnIleCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnA 283
         ::::::|||:::|||:::   :::||||||   |||||||||   :::|
     907 GGAGCTCTGTGGAAATCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAA 956

     283 snMetAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGly 299
         ||:::|||||||||||||:::::::||||||:::::::|||:::   ||||||
     957 ATTTAAATATGTCTAGAGTGGATGTATATACAACACATTCTCCTGCTGGA 1006

     300 ThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGl 316
         ||||||||||||||||:::|||||||||||||||||||||||:::      ::
    1007 ACTTCTGTGCAAAACATGTTACACTGGAGCCAGGCTGTTAAATTCCAAAA 1056

     316 uLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysA 333
         :::::::|||||||||||||||||:::::::||||||              |
    1057 GTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCATTACA 1106

     333 snGlnProThrProValArgTyrArgValArgAspMetThrValProThr 349
         |||||      |||      |||:::|||:::||||||   |||||||||
    1107 ACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACT 1156

     350 AlaMetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLy 366
         |||:::|||:::|||||||:::|||||||||:::::::      ||||||::
    1157 GCAGTCTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAA 1206

     366 sMetLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIleProG 383
         ::::|||||:::::::::|||||||||:::::|||:::::|||||||
    1207 TATCTTACTGACTCAGATCACCAACTTGGTGTTCCATGAGAGCATTCCGG 1256

     383 luTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMet 399
         |||||   |||:::||||||||||||||||||||||||||   |||:::
    1257 AATGGGAGCATCTTGACTTCATTTGGGGCCTGGATGCCCCTTGGAGGCTT 1306

     400 TyrAsnGluIleIleHisLeuMetGlnGln 409
         |||||||:::||||||:::||||||:::::::
    1307 TATAATAAAATTATTAATCTAATGAGGAAA 1336

 seq_name: gb_pr:HUMLIPCHL

 seq_documentation_block:
 LOCUS       HUMLIPCHL    2493 bp    mRNA            PRI       07-JAN-1995
 DEFINITION  Human lysosomal acid lipase/cholesteryl esterase mRNA, complete
             cds.
 ACCESSION   M74775
 VERSION     M74775.1  GI:187151
 KEYWORDS    lysosomal acid lipase/cholesteryl esterase.
 SOURCE      Homo sapiens fibroblast cDNA to mRNA.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 2493)
   AUTHORS   Anderson,R.A. and Sando,G.N.
   TITLE     Cloning and expression of cDNA encoding human lysosomal acid
             lipase/cholesteryl ester hydrolase. Similarities to gastric and
             lingual lipases
 -> JOURNAL  J. Biol. Chem. 266 (33), 22479-22484 (1991)
   MEDLINE   92042192
 FEATURES             Location/Qualifiers
      source          1. .2493
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /tissue_type="fibroblast"
                      /map="10"
      mRNA            <1. .2493
                      /gene="LIPA"
                      /note="G00-120-153"
      gene            1. .2493
                      /gene="LIPA"
      CDS             41. .1240
                      /gene="LIPA"
                      /EC_number="3.1.1.13"
                      /note="alternate initiator ATG may be at position 47"
                      /codon_start=1
                      /db_xref="GDB:G00-120-153"
                      /product="lysosomal acid lipase/cholesteryl esterase"
                      /protein_id="AAA59519.1"
                      /db_xref="GI:187152"
```

```
                     /translation="MKMRFLGLVVCLVLWPLHSEGSGGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     sig_peptide     41. .103
                     /gene="LIPA"
                     /note="alternate signal peptide may begin at position 47;
                     as predicted from von Heijne, G. (1986) Nucl. Acids Res.
                     14,4683-4690; G00-120-153"
     mat_peptide     104. .1237
                     /gene="LIPA"
                     /EC_number="3.1.1.13"
                     /note="G00-120-153"
                     /product="lysosomal acid lipase/cholesteryl esterase"
     polyA_signal    2466. .2471
                     /gene="LIPA"
                     /note="G00-120-153"
     polyA_site      2493
                     /gene="LIPA"
                     /note="G00-120-153"
BASE COUNT      692 a    484 c    516 g    801 t
ORIGIN


alignment_scores:
            Quality: 1306.00              Length:    410
              Ratio:  3.841                Gaps:      2
 Percent Similarity:  82.927   Percent Identity: 60.732

alignment_block:
 US-09-578-063-47 x HUMLIPCHL   ..

 Align seg 1/1  to: HUMLIPCHL  from: 1  to: 2493

      2 LeuGluThrLeuSerArgGlnTrpIleValSerHisArgMetGluMetTr 18
        |||||||||   :::|||||||         |||   |||||||:::|||
      8 CTCGAGACAGCGGCCCGGCAG......GACAGCTCCAGAATGAAAATGCG 51

     18 pLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAsnSerValH 35
         :::|||   |||||||:::  :::                :::::::|||
     52 GTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGCCCCTGCATTCTGAGG 101

     35 isMetProThrLys......AlaValAspProGluAlaPheMetAsnIle 49
                   |||      |||||||||||||||:::   |||||||:::
    102 GGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTG 151

     50 SerGluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAl 66
        ||||||||||||:::::   |||:::|||   |||||||||   |||
    152 AGTGAAATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGA 201

     66 aThrGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuV 83
         ||||||||||||||||||||   :::|||||||||||||:::|||
    202 GACAGAAGATGGATATATTCTGTGCCTTAACCGAATTCCTCATGGGAGGA 251

     83 alGlnProLysLysThrGlySerArgProValValLeuLeuGlnHisGly 99
         :::   :::      |||   :::||||||||||:::||||||||||||
    252 AGAACCATTCTGACAAAGGTCCCAAACCAGTTGTCTTCCTGCAACATGGC 301

    100 LeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLe 116
        |||::::::   :::|||||||||::::::||||||   |||:::|||||
    302 TTGCTGGCAGATTCTAGTAACTGGGTCACAAACCTTGCCAACAGCAGCCT 351

    116 uGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSerA 133
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    352 GGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAACAGCA 401

    133 rgGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAsp 149
        ||||||||:::||||||||||||||||||||||||||||::::::||||||
    402 GAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGAT 451

    150 GluPheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeuProAl 166
        ||||||||||||||||||||||||||||||||||:::::||||||||||
    452 GAATTCTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGC 501

    166 aValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrV 183
        |   ||||||||||||||:::|||||||||||||||:::::|||||||
    502 TTCCATTAACTTCATTCTGAATAAAACTGGCCAAGAACAAGTGTATTATG 551

    183 alGlyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMet 199
        |||||:::|||||||||||||||:::|||||||||||||||||   :::
    552 TGGGTCATTCTCAAGGCACCACTATAGGTTTTATAGCATTTTCACAGATC 601

    200 ProGluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProIleAl 216
        ||||||||||||:::::|||||||||:::||||||||||:::|||:::||
    602 CCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGCCCTGGGTCCTGTGGC 651

    216 aThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeuProA 233
        |:::|||      :::   ||||||   :::|||:::      |||||||
    652 TTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATTACCAG 701

    233 spMetMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThr 249
        ||   :::||||||   |||||||||   ||||||||||||   |||:::
    702 ATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGT 751

    250 ArgPheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAs 266
           |||||:::   |||      :::::::|||   :::||||||||||
    752 GCGTTTTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAA 801

    266 pGlnIleCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnA 283
        ::::::|||:::|||:::   :::||||||   |||||||||   :::|
    802 GGAGCTCTGTGGAAATCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAA 851

    283 snMetAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGly 299
        ||:::|||||||||||||:::::|||||:::::|||:::   ||||||
    852 ATTTAAATATGTCTAGAGTGGATGTATATACAACACATTCTCCTGCTGGA 901

    300 ThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGl 316
        ||||||||||||||||:::||||||||||||||||||||||:::      ::
    902 ACTTCTGTGCAAAACATGTTACACTGGAGCCAGGCTGTTAAATTCCAAAA 951.

    316 uLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysA 333
        :::::::|||||||||||||||||:::::|||||                |
    952 GTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCATTACA 1001

    333 snGlnProThrProValArgTyrArgValArgAspMetThrValProThr 349
        |||||      |||      |||:::|||:::||||||   |||||||||
   1002 ACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACT 1051

    350 AlaMetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLy 366
        |||:::|||:::||||||:::||||||||||:::::      ||||||::
   1052 GCAGTCTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAA 1101

    366 sMetLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIleProG 383
        ::::||||||:::::::::||||||||:::::|||:::::||||||||||
   1102 TATCTTACTGACTCAGATCACCAACTTGGTGTTCCATGAGAGCATTCCGG 1151

    383 luTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMet 399
        |||||   |||:::|||||||||||||||||||||||||||   |||:::
   1152 AATGGGAGCATCTTGACTTCATTTGGGGCCTGGATGCCCCTTGGAGGCTT 1201

    400 TyrAsnGluIleIleHisLeuMetGlnGln 409
        ||||||:::|||||||:::|||||||:::::
   1202 TATAATAAAATTATTAATCTAATGAGGAAA 1231

seq_name: gb_pr:BC012287

seq_documentation_block:
LOCUS       BC012287     2586 bp    mRNA            PRI       08-AUG-2001
DEFINITION  Homo sapiens, Similar to lipase A, lysosomal acid, cholesterol
            esterase (Wolman disease), clone MGC:5229 IMAGE:2900168, mRNA,
            complete cds.
ACCESSION   BC012287
```

```
          Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
          OH, 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2481
                     /organism="Homo sapiens"
                     /strain="HepG2"
                     /db_xref="taxon:9606"
                     /clone="clone hLAL"
                     /tissue_type="liver"
                     /cell_type="hepatoma"
                     /clone_lib="human liver HepG2 cDNA library/lambda ZAP"
     sig_peptide     34. .96
                     /gene="LAL"
     gene            34. .1233
                     /gene="LAL"
     CDS             34. .1233
                     /gene="LAL"
                     /citation=[1]
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="CAA83495.1"
                     /db_xref="GI:506431"
                     /db_xref="SPTREMBL:Q16529"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     variation       79
                     /gene="LAL"
                     /citation=[1]
                     /replace="c"
     mat_peptide     97. .1230
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     variation       100
                     /gene="LAL"
                     /citation=[1]
                     /replace="g"
     polyA_signal    2459. .2464
     polyA_site      2480
BASE COUNT      692 a    477 c    513 g    799 t
ORIGIN
```

```
  Query Match             41.6%;  Score 527.8;  DB 9;  Length 2481;
  Best Local Similarity   65.4%;  Pred. No. 7.3e-147;
  Matches  791;  Conservative    0;  Mismatches  412;  Indels    6;  Gaps    1;

Qy      39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
           ||||||| ||||| || | ||    ||||||  | | || | |   | |   || ||||
Db      30 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTC 89

Qy      99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152
           ||      |  |   | |||||||||| || ||| ||  |||||||  | |||||
Db      90 TGAGGGGTCTAGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 149

Qy     153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212
           ||| |||    |    || |  ||  ||||||||||   ||| |  || ||||||||  ||
Db     150 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 209

Qy     213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272
           ||| || |   ||||| | |||||||  ||      | | |    | | |||| |||
Db     210 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 269

Qy     273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332
           || || || || ||||| ||||||  |  | || | |||| ||||||  |  | |||||
Db     270 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 329

Qy     333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392
```

```
           |||||| ||||||||||||||||| || |||||||||||||||||||||||||||||| ||
Db     330 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 389

Qy     393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452
           |||||| |||||  |||||||||| ||||| ||||||||||  |    || ||||| ||
Db     390 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 449

Qy     453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512
           |||||||||||||||||||||||||||||  |  | ||||| || ||    || ||||| ||
Db     450 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 509

Qy     513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572
           | || | ||||| ||||| ||| |  | ||||||| ||  |||| || |||||||| ||
Db     510 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 569

Qy     573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632
           || ||||| ||||||||    || || |||||||||| | |  || |||||||| ||||||
Db     570 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 629

Qy     633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 692
           | |  ||  | ||  | ||          |  |||||  | |||||||        |
Db     630 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 689

Qy     693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752
           ||||||    | || |||| || |||||  | ||||||||||    ||||     ||
Db     690 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 749

Qy     753 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 812
           | | |     || | ||    || ||||    || || || ||  |  || | | || | ||
Db     750 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 809

Qy     813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872
           | ||   || |||||| ||||||| ||    |  ||| | || |||    |||    |||
Db     810 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 869

Qy     873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932
           |||| |  | ||  ||| |||||||| |||||||||| ||  ||||||||||||||||
Db     870 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 929

Qy     933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992
           || || |      |  | |  || |||||||||| ||    |||| |||       |
Db     930 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 989

Qy     993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052
           | ||| |||      ||||  |  ||||  || |  || |||   || || || ||| |
Db     990 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1049

Qy    1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112
           ||||  || ||||| |||||||| || ||    | |||| || ||  | ||  || |
Db    1050 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1109

Qy    1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172
           | | |||||| |  | | |||| ||| ||||| |||||||  ||  | || ||||| ||
Db    1110 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1169

Qy    1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232
           |||  |||||||  |||    |  | || ||| |||| ||  |||| |||  | |  | |
Db    1170 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1229

Qy    1233 gaccaacct 1241
           |   || ||
Db    1230 GTGAAAGCT 1238
```

```
RESULT   2
HSU08464
LOCUS       HSU08464     2481 bp    mRNA            PRI       23-JUN-1994
DEFINITION  Human lysosomal acid lipase mRNA, complete cds.
ACCESSION   U08464
VERSION     U08464.1  GI:505052
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 2481)
AUTHORS    Du,H. and Gregory,G.A.
TITLE      Structural Conservation of Putative Functional Motifs between Mouse
           and Human Lysosomal Acid Lipase
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 2481)
AUTHORS    Du,H.
TITLE      Direct Submission
JOURNAL    Submitted (08-APR-1994) Hong Du, Division of Human Gentics,
           Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
           OH 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2481
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="clone hLAL"
                     /clone_lib="HepG2 cDNA library/Lambda ZAP"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /tissue_type="liver"
     5'UTR           1. .33
     sig_peptide     34. .96
     CDS             34. .1233
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="AAB60328.1"
                     /db_xref="GI:505053"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     variation       79
                     /replace="c"
     mat_peptide     97. .1230
                     /product="lysosomal acid lipase"
     variation       100
                     /replace="g"
     3'UTR           1234. .2480
     polyA_signal    2459. .2464
     polyA_site      2480
BASE COUNT      692 a    477 c    513 g    799 t
ORIGIN
```

Query Match 41.6%; Score 527.8; DB 9; Length 2481;
Best Local Similarity 65.4%; Pred. No. 7.3e-147;
Matches 791; Conservative 0; Mismatches 412; Indels 6; Gaps 1;

```
Qy     39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
Db     30 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTC 89

Qy     99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152
Db     90 TGAGGGGTCTAGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 149

Qy    153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212
Db    150 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 209

Qy    213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272
Db    210 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 269

Qy    273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332
Db    270 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 329

Qy    333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392
Db    330 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 389

Qy    393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452
Db    390 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 449

Qy    453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512
Db    450 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 509

Qy    513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572
Db    510 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 569

Qy    573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632
Db    570 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 629

Qy    633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 692
Db    630 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 689

Qy    693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752
Db    690 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 749

Qy    753 tctcagacaacttgttatttaccttttgtggccaggtgattcttgatcagatttgtagtaa 812
Db    750 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 809

Qy    813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872
Db    810 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 869

Qy    873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932
Db    870 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 929

Qy    933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992
Db    930 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 989

Qy    993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052
Db    990 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1049

Qy   1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112
Db   1050 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1109

Qy   1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172
Db   1110 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1169

Qy   1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232
Db   1170 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1229

Qy   1233 gaccaacct 1241
Db   1230 GTGAAAGCT 1238
```

```
RESULT   3
HSLAL
LOCUS       HSLAL        2626 bp    mRNA            PRI       25-FEB-1994
DEFINITION  H.sapiens mRNA for lysosomal acid lipase.
ACCESSION   X76488
VERSION     X76488.1  GI:434305
KEYWORDS    lysosomal acid lipase.
```

GenCore version 4.5

nucleic search, using sw model

January 30, 2002, 20:27:03 ; Search time 90.11 Seconds
(without alignments)
3189.442 Million cell updates/sec

US-09-578-063-46
Score: 1269
1 atgttggaaaccttgtcaag.....gacggtgtgaggccgtattg 1269

table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

ed: 351203 seqs, 113238999 residues

number of hits satisfying chosen parameters: 702406

Num DB seq length: 0
Num DB seq length: 2000000000

t-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

tabase : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	444.2	35.0	1137	1	US-08-227-108-2	Sequence 2, Appli
	2	444.2	35.0	1137	2	US-09-073-674-2	Sequence 2, Appli
	3	444.2	35.0	1140	1	US-08-227-108-4	Sequence 4, Appli
	4	444.2	35.0	1140	2	US-09-073-674-4	Sequence 4, Appli
	5	444.2	35.0	1146	1	US-08-227-108-6	Sequence 6, Appli
	6	444.2	35.0	1146	2	US-09-073-674-6	Sequence 6, Appli
	7	444.2	35.0	1528	1	US-08-227-108-1	Sequence 1, Appli
	8	444.2	35.0	1528	2	US-09-073-674-1	Sequence 1, Appli
	9	48.2	3.8	178	1	US-08-751-782-2	Sequence 2, Appli
	10	48.2	3.8	178	2	US-08-925-171-2	Sequence 2, Appli
	11	46.4	3.7	43676	3	US-09-356-952-12	Sequence 12, Appl
	12	36.6	2.9	3271	2	US-08-852-806-1	Sequence 1, Appli
	13	36.6	2.9	3271	3	US-09-163-669-1	Sequence 1, Appli
	14	36.6	2.9	3282	1	US-08-276-852-154	Sequence 154, App
c	15	36.6	2.9	3282	1	US-08-276-852-169	Sequence 169, App
	16	36.6	2.9	3282	1	US-08-899-575-154	Sequence 154, App
c	17	36.6	2.9	3282	1	US-08-899-575-169	Sequence 169, App
	18	36.6	2.9	3282	1	US-08-899-575-154	Sequence 154, App
c	19	36.6	2.9	3282	1	US-08-899-575-169	Sequence 169, App
	20	36.6	2.9	3282	5	PCT-US95-08743-154	Sequence 154, App
c	21	36.6	2.9	3282	5	PCT-US95-08743-169	Sequence 169, App
	22	36.6	2.9	13254	1	US-08-276-852-156	Sequence 156, App
c	23	36.6	2.9	13254	1	US-08-276-852-170	Sequence 170, App
	24	36.6	2.9	13254	1	US-08-899-575-156	Sequence 156, App
c	25	36.6	2.9	13254	1	US-08-899-575-170	Sequence 170, App
	26	36.6	2.9	13254	1	US-08-899-575-156	Sequence 156, App
c	27	36.6	2.9	13254	1	US-08-899-575-170	Sequence 170, App
	28	36.6	2.9	13254	5	PCT-US95-08743-156	Sequence 156, App
c	29	36.6	2.9	13254	5	PCT-US95-08743-170	Sequence 170, App
	30	33.6	2.6	8392	1	US-08-080-255-6	Sequence 6, Appli
	31	33.6	2.6	8392	3	US-08-465-713-6	Sequence 6, Appli
	32	33.6	2.6	8392	5	PCT-US93-05857-6	Sequence 6, Appli
c	33	32.6	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	34	32	2.5	2197	1	US-08-233-005-3	Sequence 3, Appli
c	35	32	2.5	2197	1	US-08-428-943-3	Sequence 3, Appli
c	36	32	2.5	2197	3	US-09-016-649-3	Sequence 3, Appli
c	37	32	2.5	2197	5	PCT-US95-04858-3	Sequence 3, Appli
	38	32	2.5	8342	3	US-08-545-860D-63	Sequence 63, Appl
	39	32	2.5	8342	5	PCT-US94-04496-63	Sequence 63, Appl
	40	31.6	2.5	3343	4	US-08-965-762-19	Sequence 19, Appl
	41	31	2.4	3212	3	US-08-673-814-1	Sequence 1, Appli
c	42	30.4	2.4	1529	1	US-08-336-778-1	Sequence 1, Appli
c	43	30.2	2.4	4982	3	US-08-699-103B-1	Sequence 1, Appli
	44	30.2	2.4	10079	2	US-08-476-866-20	Sequence 20, Appl
c	45	30	2.4	3639	2	US-08-737-524B-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-08-227-108-2
; Sequence 2, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
US-08-227-108-2

Query Match 35.0%; Score 444.2; DB 1; Length 1137;
Best Local Similarity 62.8%; Pred. No. 1.5e-134;
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcatttttccaccatgccagagc 604
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105
```

```
Qy   1205 aaatcatccatctgatg 1221
Db   1106 AAATTGTTTCCATGATG 1122
```

```
RESULT   2
US-09-073-674-2
; Sequence 2, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1137 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1137
US-09-073-674-2
```

```
  Query Match           35.0%;  Score 444.2;  DB 2;  Length 1137;
  Best Local Similarity   62.8%;  Pred. No. 1.5e-134;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;
```

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
```

OM of: US-09-578-063-47 to: Issued_Patents_NA:* out_format : pfs

Date: Jan 31, 2002 12:05 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09578063/runat_30012002_092151_26325/app_query.fasta_1.488
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09578063_@CGN1_1_81 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-578-063-47
Query length: 423
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 90.070000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-2	+	1141.50	2461.53	1.9e-129	1137	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-2	+	1141.50	2461.53	1.9e-129	1137	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-4	+	1141.50	2461.49	1.9e-129	1140	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-4	+	1141.50	2461.49	1.9e-129	1140	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-6	+	1141.50	2461.43	1.9e-129	1146	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-6	+	1141.50	2461.43	1.9e-129	1146	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-1	+	1141.50	2457.71	3.0e-129	1528	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-1	+	1141.50	2457.71	3.0e-129	1528	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-356-952-12	+	271.50	510.09	9.2e-21	43676	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-751-782-2	+	155.00	326.10	1.6e-10	178	S
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-925-171-2	+	155.00	326.10	1.6e-10	178	S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-484-105-21	+	105.00	187.71	0.0083	1676	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-484-106-21	+	105.00	187.71	0.0083	1676	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-602-359A-24	+	102.00	187.29	0.0088	1041	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-232-519-1	+	98.50	178.16	0.0284	1167	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-956-1	+	98.50	178.16	0.0284	1167	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-115-446-1	+	92.00	153.99	0.6298	2521	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-115-446-5	+	92.00	153.96	0.6318	2526	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-119A-1	+	92.00	116.26	79.56	46899	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-080-386-3	-	91.00	153.79	0.6457	2160	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-390-000A-3	-	91.00	153.79	0.6457	2160	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-742-185-102	-	91.00	116.02	82.06	40328	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-968-542C-1	+	89.00	136.17	6.19	6027	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-755-559-2	-	85.00	140.54	3.53	2180	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-210-474-2	-	85.00	140.54	3.53	2180	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-602-359A-30	+	84.50	149.29	1.15	1017	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-759-945-1	+	84.50	137.05	5.53	2625	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-707-367-1	+	84.50	133.77	8.41	3383	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-386-607-1	+	84.00	138.47	4.61	2160	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-882-329-4	+	82.50	141.46	3.14	1329	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-183-213-4	+	82.50	141.46	3.14	1329	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-446-803-4	+	82.50	140.29	3.65	1455	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-861-837-4	+	82.50	140.29	3.65	1455	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-600-656-4	+	82.50	140.29	3.65	1455	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-170-670-9	+	82.50	140.29	3.65	1455	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-170-670-14	+	82.50	140.29	3.65	1455	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-193-068-9	+	82.50	140.29	3.65	1455	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-193-068-13	+	82.50	140.29	3.65	1455	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-183-412-9	+	82.50	140.29	3.65	1455	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-183-412-13	+	82.50	140.29	3.65	1455	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-354-191A-4	+	82.50	140.29	3.65	1455	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-600-908A-11	+	82.50	140.26	3.66	1458	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-683-838A-11	+	82.50	140.26	3.66	1458	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-483-101-1	+	82.50	122.44	35.99	5798	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-348-891A-1	+	82.50	109.43	191.04	15894	

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-2

seq_documentation_block:
; Sequence 2, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Panucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
US-08-227-108-2

alignment_scores:
Quality: 1141.50 Length: 372
Ratio: 3.624 Gaps: 3
Percent Similarity: 84.677 Percent Identity: 56.452

alignment_block:
US-09-578-063-47 x US-08-227-108-2

Align seg 1/1 to: US-08-227-108-2 from: 1 to: 1137

```
      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
         ||||||          :::||||||:::   ||||||||||||:::::||
      19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
         |   :::   ||||||||||:::|||||||||||||||:::|||||||||
      60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

      70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
```

```
     |||||||||||:::::::::|||||||||   |||      :::   :::
110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

 87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
    :::   |||   |||||||||:::::|||||||||||||||:::::::
160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
    :|||:::|||||||||||||||||||||||||||||||:::|||||||
210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
    |||||||||||||:::|||||||||:::|||||||||||||||||||:::
260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
    |||:::|||::::::          |||   |||:::   ||||||||||
310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 359

153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
    |||||||:::|||||||||||||:::::|||||||||||||:::|||:::|
360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409

170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
    ||||||||:::|||||||||||||:::|||:::::|||||||||:::|||
410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
    ||||||||||||:::|||||||||||||||||||||   |||:::|||||
460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
    |:::::::||||||   :::::::|||||||||||||:::|||||||||||:
510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
    ::::::::::::      :::|||:::::::|||:::|||:::::::::::::
560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
    |||   :::|||||||:::|||   |||   |||      :::|||:::
610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
     |||   |||:::     :::|||:::::::      :::|||   :::|
657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
    ||||||||   ::::::::::::   ||||||:::|||   |||:::|||
707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
    |||||||||   :::||||||   :::|||:::   ||||||||||||||
757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
    |||||||:::|||||||||||||||||||||:::||||||:::::::::|
807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
    ||||||||||||||||   :::::::|||:::         :::|||
857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
       |||      |||::::::   ||||||   ||||||   |||:::||
907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
    |:::|||||||:::|||   |||:::::::|||:::||||||   :::||||
957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
    ''''|::::::   ||||||||||||||||:::::::||||||   :::
```

```
1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

 386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
     |||:::|||||||||||||:::::::||||||||||:::   :::|||||||
1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

 402 uIleIleHisLeuMet 407
     ||||:::   :::|||
1107 AATTGTTTCCATGATG 1122

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-2

seq_documentation_block:
; Sequence 2, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1137 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1137
US-09-073-674-2


alignment_scores:
            Quality: 1141.50              Length:    372
              Ratio:  3.624                Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
 US-09-578-063-47 x US-09-073-674-2   ..

 Align seg 1/1  to: US-09-073-674-2  from: 1  to: 1137

      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
         ||||||          :::||||||:::   |||||||||||:::::||
      19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
```

60 / CACCA...

Query Match 51.2%; Score 1150.5; DB 1; Length 379;
Best Local Similarity 56.8%; Pred. No. 2.6e-113;
Matches 213; Conservative 65; Mismatches 94; Indels 3; Gaps 1;

```
Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| |:||| ||||||| ||||| |     | |||| |
Db       7 PT---NPEVNMNISQMISYWGYPSEKYEVVTEDGYILEVNRIPYGKKNSGNRGQRPVVFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||: |||||||||||||| |||||||: ||:|||||| |||::   | |  ||||||:
Db      64 QHGLLASASNWISNLPNNSLAFILADAGYGVWLGNSRGNTWSRRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:||:::|||||::|||.|||||:|||||| |:||::|| ::|||||:|
Db     124 DEMAKYDLPATIDFIVKETGQEKLHYVGHSQGTTIGFIAFSTNPKLAERIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLL 276
           ||||: || .|  :|  |.|.|||:| |     | : |   :| :  |: |||| : ::
Db     184 TVKYTKSLVNKLRFIPPTMFKIIFGDKIFYPHNFFDQFLATQVCSRETLNVICSNALFII 243

Qy     277 GGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPT 336
           |||:: |:|||| || :| |||||||:|||:||| ||  :||:|||  :|:   ||||
Db     244 CGFDSANLNMSRLDVYVSHNPAGTSVQNMLHWTQAVKSGNFQAFNWGSPAQNVVHFNQPT 303

Qy     337 PVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAP 396
           |  | |  | || |:|:|| |||:|:|| :|| :::|||||| |  : |:|||| ::||
Db     304 PPYYNVTAMNVPIAVWSGGNDWLADPQDVDLLLPKLSNLIYHKEILPYNHLDFIWAMNAP 363

Qy     397 HRMYNEIIHLMQQEE 411
             :||||| :| :::
Db     364 QEVYNEIISMMAKDK 378
```

RESULT 4
US-09-073-674-18
; Sequence 18, Application US/09073674
; Patent No. 5998189
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warner-Lambert Company
; STREET: 2800 Plymouth Road
; CITY: Ann Arbor
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,674
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crissey, Todd M.
; REGISTRATION NUMBER: 37,807
; REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 734 622-7530
; TELEFAX: 734 622-1553
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-073-674-18

Query Match 51.2%; Score 1150.5; DB 2; Length 379;
Best Local Similarity 56.8%; Pred. No. 2.6e-113;
Matches 213; Conservative 65; Mismatches 94; Indels 3; Gaps 1;

```
Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| |:||| ||||||| ||||| |       | |||| |
Db       7 PT---NPEVNMNISQMISYWGYPSEKYEVVTEDGYILEVNRIPYGKKNSGNRGQRPVVFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||: |||||||||||||| |||||||: ||:|||||| |||::   | |  ||||||:
Db      64 QHGLLASASNWISNLPNNSLAFILADAGYGVWLGNSRGNTWSRRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:||:::|||||::|||.|||||:|||||| |:||::|| ::|||||:|
Db     124 DEMAKYDLPATIDFIVKETGQEKLHYVGHSQGTTIGFIAFSTNPKLAERIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLL 276
           ||||: || .|  :|  |.|.|||:| |     | : |   :| :  |: |||| : ::
Db     184 TVKYTKSLVNKLRFIPPTMFKIIFGDKIFYPHNFFDQFLATQVCSRETLNVICSNALFII 243

Qy     277 GGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPT 336
           |||:: |:|||| || :| |||||||:|||:||| ||  :||:|||  :|:   ||||
Db     244 CGFDSANLNMSRLDVYVSHNPAGTSVQNMLHWTQAVKSGNFQAFNWGSPAQNVVHFNQPT 303

Qy     337 PVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAP 396
           |  | |  | || |:|:|| |||:|:|| :|| :::|||||| |  : |:|||| ::||
Db     304 PPYYNVTAMNVPIAVWSGGNDWLADPQDVDLLLPKLSNLIYHKEILPYNHLDFIWAMNAP 363

Qy     397 HRMYNEIIHLMQQEE 411
             :||||| :| :::
Db     364 QEVYNEIISMMAKDK 378
```

RESULT 5
US-08-227-108-3
; Sequence 3, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE

```
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  379 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-227-108-3


  Query Match             50.8%;  Score 1141.5;  DB 1;  Length 379;
  Best Local Similarity   56.5%;  Pred. No. 2.4e-112;
  Matches  210;  Conservative   67;  Mismatches   90;  Indels    5;  Gaps    3;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| ||||| ||||||| ::||| |    :  | |||  |
Db       7 PT---NPEVTMNISQMITYWGYPAEEYEVVTEDGYILGIDRIPYGRKNSENIGRRPVAFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |:|||||||||||| |||||||:|||:|||||| |:|::   | |  ||||||:
Db      64 QHGLLASATNWISNLPNNSLAFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:|||:||||:|::|||:|||||:|||||| |:||::|| ::|||||:|
Db     124 DEMAKYDLPATIDFILKKTGQDKLHYVGHSQGTTIGFIAFSTNPKLAKRIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLL 275
           |||: ::   | :|:|  : | :|| | | |   |  | |   :| :  :| :||| : :
Db     184 TVKYTETLLNKLMLVPSFLFKLIFGNKIF-YPHHFFDQFLATEVCSRETVDLLCSNALFI 242

Qy     276 LGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQP 335
           : ||:| |:|||  || :|  |||||||:||||||| ||: :||||||  :|:   :|
Db     243 ICGFDTMNLNMSRLDVYLSHNPAGTSVQNVLHWSQAVKSGKFQAFDWGSPVQNMMHYHQS 302

Qy     336 TPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDA 395
           |  | : || || |:| || | |::| || :|||:: ||||| || : |:|||| :||
Db     303 MPPYYNLTDMHVPIAVWNGGNDLLADPHDVDLLLSKLPNLIYHRKIPPYNHLDFIWAMDA 362

Qy     396 PHRMYNEIIHLM 407
           |  :||||: :|
Db     363 PQAVYNEIVSMM 374


RESULT   6
US-09-073-674-3
; Sequence 3, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  379 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-073-674-3


  Query Match             50.8%;  Score 1141.5;  DB 2;  Length 379;
  Best Local Similarity   56.5%;  Pred. No. 2.4e-112;
  Matches  210;  Conservative   67;  Mismatches   90;  Indels    5;  Gaps    3;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| ||||| ||||||| ::||| |    :  | |||  |
Db       7 PT---NPEVTMNISQMITYWGYPAEEYEVVTEDGYILGIDRIPYGRKNSENIGRRPVAFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |:|||||||||||| |||||||:|||:|||||| |:|::   | |  ||||||:
Db      64 QHGLLASATNWISNLPNNSLAFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:|||:||||:|::|||:|||||:|||||| |:||::|| ::|||||:|
Db     124 DEMAKYDLPATIDFILKKTGQDKLHYVGHSQGTTIGFIAFSTNPKLAKRIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLL 275
           |||: ::   | :|:|  : | :|| | | |   |  | |   :| :  :| :||| : :
Db     184 TVKYTETLLNKLMLVPSFLFKLIFGNKIF-YPHHFFDQFLATEVCSRETVDLLCSNALFI 242

Qy     276 LGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQP 335
           : ||:| |:|||  || :|  |||||||:||||||| ||: :||||||  :|:   :|
Db     243 ICGFDTMNLNMSRLDVYLSHNPAGTSVQNVLHWSQAVKSGKFQAFDWGSPVQNMMHYHQS 302

Qy     336 TPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDA 395
           |  | : || || |:| || | |::| || :|||:: ||||| || : |:|||| :||
Db     303 MPPYYNLTDMHVPIAVWNGGNDLLADPHDVDLLLSKLPNLIYHRKIPPYNHLDFIWAMDA 362

Qy     396 PHRMYNEIIHLM 407
           |  :||||: :|
Db     363 PQAVYNEIVSMM 374


RESULT   7
US-08-227-108-5
; Sequence 5, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

nucleic search, using sw model

January 30, 2002, 20:27:03 ; Search time 90.11 Seconds
(without alignments)
3189.442 Million cell updates/sec

US-09-578-063-46
score: 1269
1 atgttggaaaccttgtcaag.....gacggtgtgaggccgtattg 1269

table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

ed: 351203 seqs, 113238999 residues

number of hits satisfying chosen parameters: 702406

num DB seq length: 0
num DB seq length: 2000000000

t-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

tabase : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	444.2	35.0	1137	1	US-08-227-108-2	Sequence 2, Appli
	2	444.2	35.0	1137	2	US-09-073-674-2	Sequence 2, Appli
	3	444.2	35.0	1140	1	US-08-227-108-4	Sequence 4, Appli
	4	444.2	35.0	1140	2	US-09-073-674-4	Sequence 4, Appli
	5	444.2	35.0	1146	1	US-08-227-108-6	Sequence 6, Appli
	6	444.2	35.0	1146	2	US-09-073-674-6	Sequence 6, Appli
	7	444.2	35.0	1528	1	US-08-227-108-1	Sequence 1, Appli
	8	444.2	35.0	1528	2	US-09-073-674-1	Sequence 1, Appli
	9	48.2	3.8	178	1	US-08-751-782-2	Sequence 2, Appli
	10	48.2	3.8	178	2	US-08-925-171-2	Sequence 2, Appli
	11	46.4	3.7	43676	3	US-09-356-952-12	Sequence 12, Appl
	12	36.6	2.9	3271	2	US-08-852-806-1	Sequence 1, Appli
	13	36.6	2.9	3271	3	US-09-163-669-1	Sequence 1, Appli
	14	36.6	2.9	3282	1	US-08-276-852-154	Sequence 154, App
c	15	36.6	2.9	3282	1	US-08-276-852-169	Sequence 169, App
	16	36.6	2.9	3282	1	US-08-899-575-154	Sequence 154, App
c	17	36.6	2.9	3282	1	US-08-899-575-169	Sequence 169, App
	18	36.6	2.9	3282	1	US-08-899-575-154	Sequence 154, App
c	19	36.6	2.9	3282	1	US-08-899-575-169	Sequence 169, App
	20	36.6	2.9	3282	5	PCT-US95-08743-154	Sequence 154, App
c	21	36.6	2.9	3282	5	PCT-US95-08743-169	Sequence 169, App
	22	36.6	2.9	13254	1	US-08-276-852-156	Sequence 156, App
c	23	36.6	2.9	13254	1	US-08-276-852-170	Sequence 170, App
	24	36.6	2.9	13254	1	US-08-899-575-156	Sequence 156, App
c	25	36.6	2.9	13254	1	US-08-899-575-170	Sequence 170, App
	26	36.6	2.9	13254	1	US-08-899-575-156	Sequence 156, App
c	27	36.6	2.9	13254	1	US-08-899-575-170	Sequence 170, App
	28	36.6	2.9	13254	5	PCT-US95-08743-156	Sequence 156, App
c	29	36.6	2.9	13254	5	PCT-US95-08743-170	Sequence 170, App
	30	33.6	2.6	8392	1	US-08-080-255-6	Sequence 6, Appli
	31	33.6	2.6	8392	3	US-08-465-713-6	Sequence 6, Appli
	32	33.6	2.6	8392	5	PCT-US93-05857-6	Sequence 6, Appli
c	33	32.6	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	34	32	2.5	2197	1	US-08-233-005-3	Sequence 3, Appli
c	35	32	2.5	2197	1	US-08-428-943-3	Sequence 3, Appli
c	36	32	2.5	2197	3	US-09-016-649-3	Sequence 3, Appli
c	37	32	2.5	2197	5	PCT-US95-04858-3	Sequence 3, Appli
	38	32	2.5	8342	3	US-08-545-860D-63	Sequence 63, Appl
	39	32	2.5	8342	5	PCT-US94-04496-63	Sequence 63, Appl
	40	31.6	2.5	3343	4	US-08-965-762-19	Sequence 19, Appl
	41	31	2.4	3212	3	US-08-673-814-1	Sequence 1, Appli
c	42	30.4	2.4	1529	1	US-08-336-778-1	Sequence 1, Appli
c	43	30.2	2.4	4982	3	US-08-699-103B-1	Sequence 1, Appli
	44	30.2	2.4	10079	2	US-08-476-866-20	Sequence 20, Appl
c	45	30	2.4	3639	2	US-08-737-524B-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-08-227-108-2
; Sequence 2, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
US-08-227-108-2

Query Match 35.0%; Score 444.2; DB 1; Length 1137;
Best Local Similarity 62.8%; Pred. No. 1.5e-134;
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    |||||||||| ||| | || |||     |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          |||||||||| |  || |||||  || |||||||||     |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
             |  |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||| |||||||||||||||||| ||||||| |||||| |||||
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||| |||||||||||||||| ||| ||||| |  | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
          |     |||    |||   |  || |||||||||||||| | ||| ||||||||||  |
Db    326 TGTACTA.TCGCCCGACTCCGTCGAATTCTGGGCTT TCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||| ||    ||  |||| || ||| ||||||||| ||||| ||| |  | |
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||    | |||| |  || |  || |    |  | |||||    | |||||  |
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||      ||   |  |||     |  || | |  |  | | ||   ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
          |  ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  |||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | || ||  | || ||||||  |||||| || || ||||  |  |||  || |||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| ||         | ||  ||    | |   |||||   |  |||      ||||
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   || || |  ||| |||||  || ||  | |||| |||  |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctc ccgtatgtacaatg 1204
          ||   |    |||| |||| || |||||||   |||||| |||||     | |||||||
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105
```

Jan 31

```
Qy   1205 aaatcatccatctgatg 1221
          ||||  |    |||||
Db   1106 AAATTGTTTCCATGATG 1122
```

RESULT 2
US-09-073-674-2
; Sequence 2, Application US/09073674
; Patent No. 5998189
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warner-Lambert Company
; STREET: 2800 Plymouth Road
; CITY: Ann Arbor
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,674
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crissey, Todd M.
; REGISTRATION NUMBER: 37,807
; REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 734 622-7530
; TELEFAX: 734 622-1553
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
US-09-073-674-2

Query Match 35.0%; Score 444.2; DB 2; Length 1137;
Best Local Similarity 62.8%; Pred. No. 1.5e-134;
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    |||||||||| ||| | || |||     |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          |||||||||| |  || |||||  || |||||||||     |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
             |  |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||| |||||||||||||||||| ||||||| |||||| |||||
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
```

OM of: US-09-578-063-47 to: Issued_Patents_NA:* out_format : pfs

Date: Jan 31, 2002 12:05 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09578063/runat_30012002_092151_26325/app_query.fasta_1.488
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09578063_@CGN1_1_81 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-578-063-47
Query length: 423
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 90.070000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-2	+	1141.50	2461.53	1.9e-129	1137	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-2	+	1141.50	2461.53	1.9e-129	1137	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-4	+	1141.50	2461.49	1.9e-129	1140	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-4	+	1141.50	2461.49	1.9e-129	1140	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-6	+	1141.50	2461.43	1.9e-129	1146	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-6	+	1141.50	2461.43	1.9e-129	1146	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-1	+	1141.50	2457.71	3.0e-129	1528	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-1	+	1141.50	2457.71	3.0e-129	1528	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-356-952-12	+	271.50	510.09	9.2e-21	43676	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-751-782-2	+	155.00	326.10	1.6e-10	178	! S
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-925-171-2	+	155.00	326.10	1.6e-10	178	! S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-484-105-21	+	105.00	187.71	0.0083	1676	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-484-106-21	+	105.00	187.71	0.0083	1676	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-602-359A-24	+	102.00	187.29	0.0088	1041	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-232-519-1	+	98.50	178.16	0.0284	1167	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-956-1	+	98.50	178.16	0.0284	1167	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-115-446-1	+	92.00	153.99	0.6298	2521	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-115-446-5	+	92.00	153.96	0.6318	2526	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-119A-1	+	92.00	116.26	79.56	46899	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-080-386-3	-	91.00	153.79	0.6457	2160	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-390-000A-3	-	91.00	153.79	0.6457	2160	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-742-185-102	-	91.00	116.02	82.06	40328	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-968-542C-1	+	89.00	136.17	6.19	6027	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-755-559-2	-	85.00	140.54	3.53	2180	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-210-474-2	-	85.00	140.54	3.53	2180	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-602-359A-30	+	84.50	149.29	1.15	1017	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-759-945-1	+	84.50	137.05	5.53	2625	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-707-367-1	+	84.50	133.77	8.41	3383	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-386-607-1	+	84.00	138.47	4.61	2160	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-882-329-4	+	82.50	141.46	3.14	1329	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-183-213-4	+	82.50	141.46	3.14	1329	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-446-803-4	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-861-837-4	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-600-656-4	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-170-670-9	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-170-670-14	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-193-068-9	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-193-068-13	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-183-412-9	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-183-412-13	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-354-191A-4	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-600-908A-11	+	82.50	140.26	3.66	1458	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-683-838A-11	+	82.50	140.26	3.66	1458	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-483-101-1	+	82.50	122.44	35.99	5798	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-348-891A-1	+	82.50	109.43	191.04	15894	!

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-2

seq_documentation_block:
; Sequence 2, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
US-08-227-108-2

alignment_scores:
Quality: 1141.50 Length: 372
Ratio: 3.624 Gaps: 3
Percent Similarity: 84.677 Percent Identity: 56.452

alignment_block:
US-09-578-063-47 x US-08-227-108-2 ..

Align seg 1/1 to: US-08-227-108-2 from: 1 to: 1137

```
      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
         ||||||         :::||||||:::   |||||||||||:::::||
      19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
         |   :::   |||||||||:::|||||||||||||||||:::|||||||||
      60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

      70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
```

```
110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

 87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCCAATTCTGGGC 359

153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409

170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
```

```
1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

 386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

 402 uIleIleHisLeuMet 407
1107 AATTGTTTCCATGATG 1122
```

```
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-2

seq_documentation_block:
; Sequence 2, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd L.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1137 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1137
US-09-073-674-2


alignment_scores:
            Quality: 1141.50              Length:    372
              Ratio:  3.624                Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
 US-09-578-063-47 x US-09-073-674-2    ..

 Align seg 1/1  to: US-09-073-674-2  from: 1  to: 1137

  37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
  19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

  53 eGlnHisGlnGlyTyrPheCysGluGluTyrGluValAlaThrGluAspG 70
```

Query Match 51.2%; Score 1150.5; DB 1; Length 379;
Best Local Similarity 56.8%; Pred. No. 2.6e-113;
Matches 213; Conservative 65; Mismatches 94; Indels 3; Gaps 1;

```
Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| |:||| ||||||| ||||| |       | |||| |
Db       7 PT---NPEVNMNISQMISYWGYPSEKYEVVTEDGYILEVNRIPYGKKNSGNRGQRPVVFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |||||||||||||| ||||||| ||:|||||| |||::   | |  ||||||:
Db      64 QHGLLASASNWISNLPNNSLAFILADAGYGVWLGNSRGNTWSRRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:||:::|||||::|||:|||||:||||||| |:||::|| ::|||||:|
Db     124 DEMAKYDLPATIDFIVKETGQEKLHYVGHSQGTTIGFIAFSTNPKLAERIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLL 276
           |||: ||  .|   :|  | | :|| | |     | : |   :| :  |: |||| : ::
Db     184 TVKYTKSLVNKLRFIPPTMFKIIFGDKIFYPHNFFDQFLATQVCSRETLNVICSNALFII 243

Qy     277 GGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPT 336
           ||:: |:|||  || :|  |||||||:|||:||| ||  :||:|||  :|:   ||||
Db     244 CGFDSANLNMSRLDVYVSHNPAGTSVQNMLHWTQAVKSG FQAFNWGSPAQNVVHFNQPT 303

Qy     337 PVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAP 396
           |  | |  | || |:|:|| |||::|:|| :|| :::|||||| |  : |:|||| ::||
Db     304 PPYYNVTAMNVPIAVWSGGNDWLADPQDVDLLLPKLSNLIYHKEILPYNHLDFIWAMNAP 363

Qy     397 HRMYNEIIHLMQQEE 411
           :|||||  :| :::
Db     364 QEVYNEIISMMAKDK 378
```

RESULT 4
US-09-073-674-18
; Sequence 18, Application US/09073674
; Patent No. 5998189
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warner-Lambert Company
; STREET: 2800 Plymouth Road
; CITY: Ann Arbor
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,674
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crissey, Todd M.
; REGISTRATION NUMBER: 37,807
; REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 734 622-7530
; TELEFAX: 734 622-1553
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-073-674-18

Query Match 51.2%; Score 1150.5; DB 2; Length 379;
Best Local Similarity 56.8%; Pred. No. 2.6e-113;
Matches 213; Conservative 65; Mismatches 94; Indels 3; Gaps 1;

```
Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| |:||| ||||||| ||||| |       | |||| |
Db       7 PT---NPEVNMNISQMISYWGYPSEKYEVVTEDGYILEVNRIPYGKKNSGNRGQRPVVFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |||||||||||| | ||||||| ||:|||||| |||::   | |  ||||||:
Db      64 QHGLLASASNWISNLPNNSLAFILADAGYGVWLGNSRGNTWSRRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:||:::|||||::|||:|||||:||||||| |:||::|| ::|||||:|
Db     124 DEMAKYDLPATIDFIVKETGQEKLHYVGHSQGTTIGFIAFSTNPKLAERIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLL 276
           |||: ||   |  :|  | | :|| | |     | : |   :| :  |: |||| : ::
Db     184 TVKYTKSLVNKLRFIPPTMFKIIFGDKIFYPHNFFDQFLATQVCSRETLNVICSNALFII 243

Qy     277 GGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPT 336
           ||:: |:|||  || :|  |||||||:|||:||| ||  :||:|||  :|:   ||||
Db     244 CGFDSANLNMSRLDVYVSHNPAGTSVQNMLHWTQAVKSGNFQAFNWGSPAQNVVHFNQPT 303

Qy     337 PVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAP 396
           |  | |  | || |:|:|| |||::|:|| :|| :::|||||| |  : |:|||| ::||
Db     304 PPYYNVTAMNVPIAVWSGGNDWLADPQDVDLLLPKLSNLIYHKEILPYNHLDFIWAMNAP 363

Qy     397 HRMYNEIIHLMQQEE 411
           :|||||  :| :::
Db     364 QEVYNEIISMMAKDK 378
```

RESULT 5
US-08-227-108-3
; Sequence 3, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE

```
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  379 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-227-108-3


  Query Match             50.8%;  Score 1141.5;  DB 1;  Length 379;
  Best Local Similarity   56.5%;  Pred. No. 2.4e-112;
  Matches  210;  Conservative   67;  Mismatches   90;  Indels    5;  Gaps    3;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  |||||::| : ||| ||||| ||||||| ::||| |    :  | |||  |
Db       7 PT---NPEVTMNISQMITYWGYPAEEYEVVTEDGYILGIDRIPYGRKNSENIGRRPVAFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |:||||||||||| |||||||:|||:|||||| |:|::   | |  ||||||:
Db      64 QHGLLASATNWISNLPNNSLAFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:|||:||||:|::|||:|||||:||||||| |:||::|| ::|||||:|
Db     124 DEMAKYDLPATIDFILKKTGQDKLHYVGHSQGTTIGFIAFSTNPKLAKRIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLL 275
           |||: ::   | :|:|  : | :|| | | |   | | |   :| :  :| :||| : :
Db     184 TVKYTETLLNKLMLVPSFLFKLIFGNKIF-YPHHFFDQFLATEVCSRETVDLLCSNALFI 242

Qy     276 LGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQP 335
           : ||:| |:|||| || :|  |||||||:||||||| ||: :|||||  :|:   :|
Db     243 ICGFDTMNLNMSRLDVYLSHNPAGTSVQNVLHWSQAVKSGKFQAFDWGSPVQNMMHYHQS 302

Qy     336 TPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDA 395
           |  | : || || |:| || | |::| || :|||:: |||||: || : |:|||| :||
Db     303 MPPYYNLTDMHVPIAVWNGGNDLLADPHDVDLLLSKLPNLIYHRKIPPYNHLDFIWAMDA 362

Qy     396 PHRMYNEIIHLM 407
           |  :||||: :|
Db     363 PQAVYNEIVSMM 374


RESULT   6
US-09-073-674-3
; Sequence 3, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  379 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-073-674-3


  Query Match             50.8%;  Score 1141.5;  DB 2;  Length 379;
  Best Local Similarity   56.5%;  Pred. No. 2.4e-112;
  Matches  210;  Conservative   67;  Mismatches   90;  Indels    5;  Gaps    3;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  |||||::| : ||| ||||| ||||||| ::||| |    :  | |||  |
Db       7 PT---NPEVTMNISQMITYWGYPAEEYEVVTEDGYILGIDRIPYGRKNSENIGRRPVAFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |:||||||||||| |||||||:|||:|||||| |:|::   | |  ||||||:
Db      64 QHGLLASATNWISNLPNNSLAFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:|||:||||:|::|||:|||||:||||||| |:||::|| ::|||||:|
Db     124 DEMAKYDLPATIDFILKKTGQDKLHYVGHSQGTTIGFIAFSTNPKLAKRIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLL 275
           |||: ::   | :|:|  : | :|| | | |   | | |   :| :  :| :||| : :
Db     184 TVKYTETLLNKLMLVPSFLFKLIFGNKIF-YPHHFFDQFLATEVCSRETVDLLCSNALFI 242

Qy     276 LGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQP 335
           : ||:| |:|||| || :|  |||||||:||||||| ||: :|||||  :|:   :|
Db     243 ICGFDTMNLNMSRLDVYLSHNPAGTSVQNVLHWSQAVKSGKFQAFDWGSPVQNMMHYHQS 302

Qy     336 TPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDA 395
           |  | : || || |:| || | |::| || :|||:: |||||: || : |:|||| :||
Db     303 MPPYYNLTDMHVPIAVWNGGNDLLADPHDVDLLLSKLPNLIYHRKIPPYNHLDFIWAMDA 362

Qy     396 PHRMYNEIIHLM 407
           |  :||||: :|
Db     363 PQAVYNEIVSMM 374


RESULT   7
US-08-227-108-5
; Sequence 5, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
```

```
OM of: US-09-578-063-47 to: GenEmbl:*   out_format : pfs

Date: Jan 31, 2002 12:03 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09578063/runat_30012002_092151_26310/app_query.fasta_1.488
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09578063_@CGN1_1_5922 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-578-063-47
Query length: 423
Database: GenEmbl:*
Database sequences: 1472140
Database length: -341344837
Search time (sec): 1647.070000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_pr:HSLAL             + 1308.00 2075.39  2.5e-107  2626  ! X76488 H.sapiens mRNA for lysc
gb_pr:HUMLIPCHL         + 1306.00 2072.67  3.6e-107  2493  ! M74775 Human lysosomal acid li
gb_pr:BC012287          + 1303.00 2067.52  6.9e-107  2586  ! BC012287 Homo sapiens, Similar
gb_pr:HSLYACLY          + 1298.00 2059.89  1.8e-106  2481  ! Z31690 H.sapiens (HepG2) LAL m
gb_pr:HSU08464          + 1298.00 2059.89  1.8e-106  2481  ! U08464 Human lysosomal acid li
gb_ro:S81497            + 1220.50 1933.43  2.0e-99   3144  ! S81497 lysosomal acid lipase=i
gb_ro:MMLYACLY          + 1206.00 1912.87  2.9e-98   2358  ! Z31689 M.musculus (C57 Black/6
gb_pat:A26689           + 1167.00 1855.35  4.6e-95   1378  ! A26689 Precursor of rabit gast
gb_pat:A26690           - 1167.00 1855.35  4.6e-95   1378  ! A26690 Precursor of rabit gast
gb_pat:A01157           + 1166.50 1854.83  4.9e-95   1336  ! A01157 R.norvegicus mRNA for p
gb_ro:RNLIP             + 1166.50 1854.70  5.0e-95   1355  ! X02309 Rat mRNA for lingual li
gb_om:CFLIPASE          + 1154.00 1832.82  8.2e-94   1651  ! Y13899 Canis familiaris mRNA f
gb_pat:AR039022         + 1141.50 1816.26  6.9e-93   1137  ! AR039022 Sequence 2 from paten
gb_pat:AR092633         + 1141.50 1816.26  6.9e-93   1137  ! AR092633 Sequence 2 from paten
gb_pat:AR039023         + 1141.50 1816.23  6.9e-93   1140  ! AR039023 Sequence 4 from paten
gb_pat:AR092634         + 1141.50 1816.23  6.9e-93   1140  ! AR092634 Sequence 4 from paten
gb_pat:AR039024         + 1141.50 1816.18  6.9e-93   1146  ! AR039024 Sequence 6 from paten
gb_pat:AR092635         + 1141.50 1816.18  6.9e-93   1146  ! AR092635 Sequence 6 from paten
gb_pat:A57760           + 1141.50 1815.77  7.3e-93   1198  ! A57760 Sequence 5 from Patent
gb_pat:A57756           + 1141.50 1813.50  9.8e-93   1528  ! A57756 Sequence 1 from Patent
gb_pat:AR039021         + 1141.50 1813.50  9.8e-93   1528  ! AR039021 Sequence 1 from paten
gb_pat:AR092632         + 1141.50 1813.50  9.8e-93   1528  ! AR092632 Sequence 1 from paten
gb_om:BOVMPE            + 1139.00 1810.20  1.5e-92   1417  ! L26319 Bos taurus pregastric e
gb_pat:A39301           + 1132.50 1801.83  4.4e-92   1137  ! A39301 Sequence 2 from Patent
gb_pat:A39303           + 1132.50 1801.80  4.4e-92   1140  ! A39303 Sequence 4 from Patent
gb_pat:A39305           + 1132.50 1801.75  4.4e-92   1146  ! A39305 Sequence 6 from Patent
gb_pat:A39300           + 1132.50 1799.07  6.2e-92   1528  ! A39300 Sequence 1 from Patent
gb_pr:HSGLR             + 1124.00 1786.50  3.1e-91   1365  ! X05997 Human mRNA for gastric
gb_pat:A01046           + 1124.00 1786.48  3.1e-91   1367  ! A01046 H.sapiens mRNA for huma
gb_pat:A12714           + 1124.00 1786.48  3.1e-91   1367  ! A12714 Pregastric lipase. 8/19
gb_pat:A57758           + 1003.00 1594.97  1.5e-80   1048  ! A57758 Sequence 3 from Patent
gb_pat:AX068259         +  695.00 1099.12  6.0e-53   1308  ! AX068259 Sequence 21 from Pate
gb_pat:AX068249         +  691.00 1098.54  6.5e-53   699   ! AX068249 Sequence 11 from Paten
gb_pr:AL353113          +  675.00 1034.41  2.4e-49   43456  ! AL353113 Human DNA sequence f
gb_htg:AC019353         +  675.00 1022.94  1.1e-48   148720  ! AC019353 Homo sapiens clone
gb_htg:AC015461         -  675.00 1022.37  1.1e-48   158195  ! AC015461 Homo sapiens clone
gb_htg:AC011277         +  675.00 1021.54  1.3e-48   172834  ! AC011277 Homo sapiens clone
gb_in:DMLIPASE3         +  665.00 1051.74  2.6e-50   1211  ! Y14367 Drosophila melanogaster
gb_in:CEF46B6           +  658.00 1008.75  6.5e-48   36616  ! Z70780 Caenorhabditis elegans
gb_in:CBRG01A11         +  638.50  976.94  3.9e-46   38822  ! AC084473 Caenorhabditis brigg
gb_in:U64849            +  637.00  974.99  5.0e-46   36982  ! U64849 Caenorhabditis elegans
gb_in:AC004766          +  636.00  964.70  1.9e-45   93932  ! AC004766 Drosophila melano
gb_in:AC008342          -  636.00  958.90  3.9e-45   175007  ! AC008342 Drosophila melan
gb_htg:AC017903         -  636.00  957.13  4.9e-45   211784  ! AC017903 Drosophila melan
gb_in:AE003811          -  636.00  955.23  6.3e-45   259445  ! AE003811 Drosophila melan

seq_name: gb_pr:HSLAL

seq_documentation_block:
LOCUS       HSLAL        2626 bp    mRNA            PRI       25-FEB-1994
DEFINITION  H.sapiens mRNA for lysosomal acid lipase.
ACCESSION   X76488
VERSION     X76488.1  GI:434305
KEYWORDS    lysosomal acid lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2626)
  AUTHORS   Ameis,D., Merkel,M., Eckerskorn,C. and Greten,H.
  TITLE     Purification, characterization and molecular cloning of human
            hepatic lysosomal acid lipase
  JOURNAL   Eur. J. Biochem. 219 (3), 905-914 (1994)
  MEDLINE   94155897
REFERENCE   2  (bases 1 to 2626)
  AUTHORS   Ameis,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-1993) D. Ameis, Medical Department, University
            Hospital Eppendorf, Martinistrasse 52, 20246 Hamburg, FRG
FEATURES             Location/Qualifiers
     source          1. .2626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /tissue_type="liver"
     gene            146. .1345
                     /gene="LAL/LIPA"
     CDS             146. .1345
                     /gene="LAL/LIPA"
                     /EC_number="3.1.1.13"
                     /note="sterol esterase"
                     /codon_start=1
                     /evidence=experimental
                     /product="lysosomal acid lipase"
                     /protein_id="CAA54026.1"
                     /db_xref="GI:434306"
                     /db_xref="SWISS-PROT:P38571"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSGGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     sig_peptide     146. .226
                     /gene="LAL/LIPA"
                     /evidence=experimental
     misc_feature    227. .373
                     /gene="LAL/LIPA"
                     /note="propeptide"
                     /evidence=experimental
     mat_peptide     374. .1342
                     /gene="LAL/LIPA"
                     /evidence=experimental
                     /product="lysosomal acid lipase"
BASE COUNT      741 a    527 c    547 g    811 t
ORIGIN


alignment_scores:
           Quality: 1308.00              Length:    410
             Ratio:  3.836                 Gaps:      2
 Percent Similarity:  83.171   Percent Identity: 60.732
```

```
alignment_block:
 US-09-578-063-47 x HSLAL   ..

 Align seg 1/1  to: HSLAL  from: 1  to: 2626

       2 LeuGluThrLeuSerArgGlnTrpIleValSerHisArgMetGluMetTr 18
       113 CTCGAGACAGCGGCCCGGCAG......GACAGCTCCAGAATGAAAATGCG 156

       18 pLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAsnSerValH 35
       157 GTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTCTGAGG 206

       35 isMetProThrLys......AlaValAspProGluAlaPheMetAsnIle 49
       207 GGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTG 256

       50 SerGluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAl 66
       257 AGTGAAATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGA 306

       66 aThrGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuV 83
       307 GACAGAAGATGGATATATTCTGTGCCTTAACCGAATTCCTCATGGGAGGA 356

       83 alGlnProLysLysThrGlySerArgProValValLeuLeuGlnHisGly 99
       357 AGAACCATTCTGACAAAGGTCCCAAACCAGTTGTCTTCCTGCAACATGGC 406

       100 LeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLe 116
       407 TTGCTGGCAGATTCTAGTAACTGGGTCACAAACCTTGCCAACAGCAGCCT 456

       116 uGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSerA 133
       457 GGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAACAGCA 506

       133 rgGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAsp 149
       507 GAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGAT 556

       150 GluPheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeuProAl 166
       557 GAATTCTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGC 606

       166 aValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrV 183
       607 TTCCATTAACTTCATTCTGAATAAAACTGGCCAAGAACAAGTGTATTATG 656

       183 alGlyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMet 199
       657 TGGGTCATTCTCAAGGCACCACTATAGGTTTTATAGCATTTTCACAGATC 706

       200 ProGluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProIleAl 216
       707 CCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGCCCTGGGTCCTGTGGC 756

       216 aThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeuProA 233
       757 TTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATTACCAG 806

       233 spMetMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThr 249
       807 ATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGT 856

       250 ArgPheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAs 266
       857 GCGTTTTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAA 906

       266 pGlnIleCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnA 283
       907 GGAGCTCTGTGGAAATCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAA 956

       283 snMetAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGly 299
       957 ATTTAAATATGTCTAGAGTGGATGTATATACAACACATTCTCCTGCTGGA 1006

       300 ThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGl 316
       1007 ACTTCTGTGCAAAACATGTTACACTGGAGCCAGGCTGTTAAATTCCAAAA 1056

       316 uLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysA 333
       1057 GTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCATTACA 1106

       333 snGlnProThrProValArgTyrArgValArgAspMetThrValProThr 349
       1107 ACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACT 1156

       350 AlaMetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLy 366
       1157 GCAGTCTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAA 1206

       366 sMetLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIleProG 383
       1207 TATCTTACTGACTCAGATCACCAACTTGGTGTTCCATGAGAGCATTCCGG 1256

       383 luTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMet 399
       1257 AATGGGAGCATCTTGACTTCATTTGGGGCCTGGATGCCCCTTGGAGGCTT 1306

       400 TyrAsnGluIleIleHisLeuMetGlnGln 409
       1307 TATAATAAAATTATTAATCTAATGAGGAAA 1336

seq_name: gb_pr:HUMLIPCHL

seq_documentation_block:
LOCUS       HUMLIPCHL    2493 bp    mRNA            PRI       07-JAN-1995
DEFINITION  Human lysosomal acid lipase/cholesteryl esterase mRNA, complete
            cds.
ACCESSION   M74775
VERSION     M74775.1  GI:187151
KEYWORDS    lysosomal acid lipase/cholesteryl esterase.
SOURCE      Homo sapiens fibroblast cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2493)
  AUTHORS   Anderson,R.A. and Sando,G.N.
  TITLE     Cloning and expression of cDNA encoding human lysosomal acid
            lipase/cholesteryl ester hydrolase. Similarities to gastric and
            lingual lipases
  JOURNAL   J. Biol. Chem. 266 (33), 22479-22484 (1991)
  MEDLINE   92042192
FEATURES             Location/Qualifiers
     source          1. .2493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="fibroblast"
                     /map="10"
     mRNA            <1. .2493
                     /gene="LIPA"
                     /note="G00-120-153"
     gene            1. .2493
                     /gene="LIPA"
     CDS             41. .1240
                     /gene="LIPA"
                     /EC_number="3.1.1.13"
                     /note="alternate initiator ATG may be at position 47"
                     /codon_start=1
                     /db_xref="GDB:G00-120-153"
                     /product="lysosomal acid lipase/cholesteryl esterase"
                     /protein_id="AAA59519.1"
                     /db_xref="GI:187152".
```

```
                     /translation="MKMRFLGLVVCLVLWPLHSEGSGGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     sig_peptide     41. .103
                     /gene="LIPA"
                     /note="alternate signal peptide may begin at position 47;
                     as predicted from von Heijne, G. (1986) Nucl. Acids Res.
                     14,4683-4690; G00-120-153"
     mat_peptide     104. .1237
                     /gene="LIPA"
                     /EC_number="3.1.1.13"
                     /note="G00-120-153"
                     /product="lysosomal acid lipase/cholesteryl esterase"
     polyA_signal    2466. .2471
                     /gene="LIPA"
                     /note="G00-120-153"
     polyA_site      2493
                     /gene="LIPA"
                     /note="G00-120-153"
BASE COUNT      692 a    484 c    516 g    801 t
ORIGIN


alignment_scores:
            Quality: 1306.00          Length:    410
              Ratio:  3.841             Gaps:      2
 Percent Similarity:  82.927   Percent Identity: 60.732

alignment_block:
 US-09-578-063-47 x HUMLIPCHL   ..

 Align seg 1/1  to: HUMLIPCHL  from: 1  to: 2493

       2 LeuGluThrLeuSerArgGlnTrpIleValSerHisArgMetGluMetTr 18
         |||||||||   :::||||||         |||   ||||||:::|||
       8 CTCGAGACAGCGGCCCGGCAG......GACAGCTCCAGAATGAAAATGCG 51

      18 pLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAsnSerValH 35
         :::|||   ||||||:::   :::             :::::|||
      52 GTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGCCCCTGCATTCTGAGG 101

      35 isMetProThrLys......AlaValAspProGluAlaPheMetAsnIle 49
                    |||      ||||||||||||||:::   ||||||:::
     102 GGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTG 151

      50 SerGluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAl 66
         ||||||||||||:::::   |||:::|||   |||||||||   |||
     152 AGTGAAATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGA 201

      66 aThrGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuV 83
         ||||||||||||||||||||   :::|||||||||||||:::|||
     202 GACAGAAGATGGATATATTCTGTGCCTTAACCGAATTCCTCATGGGAGGA 251

      83 alGlnProLysLysThrGlySerArgProValValLeuLeuGlnHisGly 99
         :::   :::      |||   :::|||||||||:::||||||||||||
     252 AGAACCATTCTGACAAAGGTCCCAAACCAGTTGTCTTCCTGCAACATGGC 301

     100 LeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLe 116
         |||::::::   :::||||||||||:::::::||||||   |||:::||||||
     302 TTGCTGGCAGATTCTAGTAACTGGGTCACAAACCTTGCCAACAGCAGCCT 351

     116 uGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSerA 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 GGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAACAGCA 401

     133 rgGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAsp 149
         ||||||||:::||||||||||||||||||||||||||:::::::||||||
     402 GAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGAT 451

     150 GluPheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeuProAl 166
         ||||||||||||||||||||||||||||||||||:::::|||||||||||
     452 GAATTCTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGC 501

     166 aValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrV 183
         |   |||||||||||||||:::|||||||||||||||||:::::|||||||
     502 TTCCATTAACTTCATTCTGAATAAAACTGGCCAAGAACAAGTGTATTATG 551

     183 alGlyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMet 199
         |||||:::|||||||||||||||||:::|||||||||||||||||||   :::
     552 TGGGTCATTCTCAAGGCACCACTATAGGTTTTATAGCATTTTCACAGATC 601

     200 ProGluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProIleAl 216
         ||||||||||||:::::|||||||||:::|||||||||:::|||:::||
     602 CCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGCCCTGGGTCCTGTGGC 651

     216 aThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeuProA 233
         |:::|||      :::   ||||||   :::|||:::      |||||||
     652 TTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATTACCAG 701

     233 spMetMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThr 249
         ||   :::||||||   |||||||||   ||||||||||||   |||:::
     702 ATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGT 751

     250 ArgPheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAs 266
            ||||||:::   |||      :::::::|||   :::|||||||||
     752 GCGTTTTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAA 801

     266 pGlnIleCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnA 283
         :::::::|||:::|||:::   :::||||||   |||||||||   :::|
     802 GGAGCTCTGTGGAAATCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAA 851

     283 snMetAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGly 299
         ||:::|||||||||||||:::::|||||:::::::||:::   ||||||
     852 ATTTAAATATGTCTAGAGTGGATGTATATACAACACATTCTCCTGCTGGA 901

     300 ThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGl 316
         ||||||||||||||:::|||||||||||||||||||||:::      ::
     902 ACTTCTGTGCAAAACATGTTACACTGGAGCCAGGCTGTTAAATTCCAAAA 951

     316 uLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysA 333
         :::::::|||||||||||||||||||:::::|||||          |
     952 GTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCATTACA 1001

     333 snGlnProThrProValArgTyrArgValArgAspMetThrValProThr 349
         |||||      |||      |||:::|||:::||||||   |||||||||
    1002 ACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACT 1051

     350 AlaMetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLy 366
         |||:::|||:::|||||||:::||||||||||:::::      ||||||::
    1052 GCAGTCTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAA 1101

     366 sMetLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIleProG 383
         ::::|||||||:::::::::|||||||||:::::|||:::::::||||||
    1102 TATCTTACTGACTCAGATCACCAACTTGGTGTTCCATGAGAGCATTCCGG 1151

     383 luTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMet 399
         |||||   |||:::||||||||||||||||||||||||||   |||:::
    1152 AATGGGAGCATCTTGACTTCATTTGGGGCCTGGATGCCCCTTGGAGGCTT 1201

     400 TyrAsnGluIleIleHisLeuMetGlnGln 409
         ||||||:::|||||||:::||||||::::::
    1202 TATAATAAAATTATTAATCTAATGAGGAAA 1231

seq_name: gb_pr:BC012287

seq_documentation_block:
LOCUS       BC012287     2586 bp    mRNA            PRI       08-AUG-2001
DEFINITION  Homo sapiens, Similar to lipase A, lysosomal acid, cholesterol
            esterase (Wolman disease), clone MGC:5229 IMAGE:2900168, mRNA,
            complete cds.
ACCESSION   BC012287
```

```
VERSION     BC012287.1  GI:15126726
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2586)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 3 Row: c Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 434305.
FEATURES             Location/Qualifiers
     source          1. .2586
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:5229 IMAGE:2900168"
                     /tissue_type="Placenta, choriocarcinoma"
                     /clone_lib="NIH_MGC_10"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             101. .1300
                     /codon_start=1
                     /product="Similar to lipase A, lysosomal acid, cholesterol
                     esterase (Wolman disease)"
                     /protein_id="AAH12287.1"
                     /db_xref="GI:15126727"
                     /translation="MKMRFLGLVVCLVLWPLHSEGSGGKLTALDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
BASE COUNT      736 a    511 c    531 g    808 t
ORIGIN


alignment_scores:
            Quality: 1303.00              Length:    410
              Ratio:  3.832                Gaps:      2
 Percent Similarity:  82.927   Percent Identity: 60.488

alignment_block:
 US-09-578-063-47 x BC012287   ..

 Align seg 1/1  to: BC012287  from: 1  to: 2586

       2 LeuGluThrLeuSerArgGlnTrpIleValSerHisArgMetGluMetTr 18
         |||||||||   :::|||||||         |||   |||||||:::|||
      68 CTCGAGACAGCGGCCCGGCAG......GACAGCTCCAGAATGAAAATGCG 111

      18 pLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAsnSerValH 35
         :::|||   ||||||:::   :::             :::::|||
     112 GTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGCCCCTGCATTCTGAGG 161

      35 isMetProThrLys......AlaValAspProGluAlaPheMetAsnIle 49
                 |||      |||:::||||||||||:::   ||||||:::
     162 GGTCTGGAGGGAAACTGACAGCTTTGGATCCTGAAACAAACATGAATGTG 211

      50 SerGluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAl 66
         ||||||||||||::::::   |||:::|||   |||||||||   |||
     212 AGTGAAATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGA 261

      66 aThrGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuV 83
         ||||||||||||||||||||   :::|||||||||||::|||
     262 GACAGAAGATGGATATATTCTGTGCCTTAACCGAATTCCTCATGGGAGGA 311

      83 alGlnProLysLysThrGlySerArgProValValLeuLeuGlnHisGly 99
         :::   :::      |||   :::||||||||||:::||||||||||||
     312 AGAACCATTCTGACAAAGGTCCCAAACCAGTTGTCTTCCTGCAACATGGC 361

     100 LeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLe 116
         |||::::::   :::||||||||||:::::::||||||   |||:::||||
     362 TTGCTGGCAGATTCTAGTAACTGGGTCACAAACCTTGCCAACAGCAGCCT 411

     116 uGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSerA 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     412 GGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAACAGCA 461

     133 rgGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAsp 149
         ||||||||:::||||||||||||||||||||||||||||:::::||||||
     462 GAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGAT 511

     150 GluPheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeuProAl 166
         ||||||||||||||||||||||||||||||||:::::|||||||||||
     512 GAATTCTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGC 561

     166 aValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrV 183
         |   ||||||||||||||:::|||||||||||||||:::::|||||||
     562 TTCCATTAACTTCATTCTGAATAAAACTGGCCAAGAACAAGTGTATTATG 611

     183 alGlyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMet 199
         |||||:::|||||||||||||||:::|||||||||||||||||   :::
     612 TGGGTCATTCTCAAGGCACCACTATAGGTTTTATAGCATTTTCAAAGATC 661

     200 ProGluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaP   leAl 216
         ||||||||||||:::::|||||||||::|||||||||:::|||:::||
     662 CCTGAGCTGGCTAAAAGGATTAAAATGTTTTTGCCCTGGGTCC   GGC 711

     216 aThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeuProA 233
         |:::|||      :::   ||||||   :::|||:::      |||||||
     712 TTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATTACCAG 761

     233 spMetMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThr 249
         ||   :::|||||||   |||||||||   ||||||||||||   |||:::
     762 ATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGT 811

     250 ArgPheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAs 266
         |||||||:::   |||      :::::|||   :::||||||||||
     812 GCGTTTTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAA 861

     266 pGlnIleCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnA 283
         ::::::|||:::|||:::   :::||||||   |||||||||   :::|
     862 GGAGCTCTGTGGAAATCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAA 911

     283 snMetAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGly 299
         ||:::|||||||||||||:::::|||||:::::|||:::   ||||||
     912 ATTTAAATATGTCTAGAGTGGATGTATATACAACACATTCTCCTGCTGGA 961

     300 ThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGl 316
         ||||||||||||||:::|||||||||||||||||||||||:::      ::
     962 ACTTCTGTGCAAAACATGTTACACTGGAGCCAGGCTGTTAAATTCCAAAA 1011
```

```
     316 uLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysA 333
         :::::::||||||||||||||||||||:::::|||||              |
    1012 GTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCATTACA 1061

     333 snGlnProThrProValArgTyrArgValArgAspMetThrValProThr 349
         |||||      |||      |||:::|||:::||||||   |||||||||
    1062 ACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACT 1111

     350 AlaMetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLy 366
         |||:::|||:::|||||||:::|||||||||:::::      ||||||::
    1112 GCAGTCTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAA 1161

     366 sMetLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIleProG 383
         ::::||||||:::::::::|||||||||:::::|||:::::|||||||
    1162 TATCTTACTGACTCAGATCACCAACTTGGTGTTCCATGAGAGCATTCCGG 1211

     383 luTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMet 399
         |||||   |||:::||||||||||||||||||||||||||   |||:::
    1212 AATGGGAGCATCTTGACTTCATTTGGGGCCTGGATGCCCCTTGGAGGCTT 1261

     400 TyrAsnGluIleIleHisLeuMetGlnGln 409
         ||||||:::||||||:::||||||::::::
    1262 TATAATAAAATTATTAATCTAATGAGGAAA 1291

seq_name: gb_pr:HSLYACLY

seq_documentation_block:
LOCUS       HSLYACLY     2481 bp    mRNA            PRI       17-FEB-1997
DEFINITION  H.sapiens (HepG2) LAL mRNA for lysosomal acid lipase.
ACCESSION   Z31690
VERSION     Z31690.1  GI:506430
KEYWORDS    LAL; lipase; lysosomal acid lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2481)
  AUTHORS   Du,H. and Gregory,G.A.
  TITLE     Structural Conservation of Putative Functional Motifs between Mouse
            and Human Lysosomal Acid Lipase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2481)
  AUTHORS   Du,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1994) Hong Du, Division of Human Genetics,
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH, 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2481
                     /organism="Homo sapiens"
                     /strain="HepG2"
                     /db_xref="taxon:9606"
                     /clone="clone hLAL"
                     /tissue_type="liver"
                     /cell_type="hepatoma"
                     /clone_lib="human liver HepG2 cDNA library/lambda ZAP"
     sig_peptide     34. .96
                     /gene="LAL"
     gene            34. .1233
                     /gene="LAL"
     CDS             34. .1233
                     /gene="LAL"
                     /citation=[1]
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="CAA83495.1"
                     /db_xref="GI:506431"
                     /db_xref="SPTREMBL:Q16529"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     variation       79
                     /gene="LAL"
                     /citation=[1]
                     /replace="c"
     mat_peptide     97. .1230
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     variation       100
                     /gene="LAL"
                     /citation=[1]
                     /replace="g"
     polyA_signal    2459. .2464
     polyA_site      2480
BASE COUNT      692 a    477 c    513 g    799 t
ORIGIN


alignment_scores:
           Quality: 1298.00              Length:    400
             Ratio:  3.875                 Gaps:      1
 Percent Similarity:  83.750   Percent Identity: 61.000

alignment_block:
 US-09-578-063-47 x HSLYACLY  ..

 Align seg 1/1  to: HSLYACLY  from: 1  to: 2481

      12 SerHisArgMetGluMetTrpLeuLeuIleLeuValAlaTyrMetPheGl 28
         |||  ||||||:::|||   :::|||   |||||:::   :::
      25 AGCTCCAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCT 74

      28 nArgAsnValAsnSerValHisMetProThrLys......AlaValAspP 43
            :::::::::|||              |||      ||||||||||
      75 CTGGACCCTGCATTCTGAGGGGTCTAGAGGGAAACTGACAGCTGTGGATC 124

      43 roGluAlaPheMetAsnIleSerGluIleIleGlnHisGlnGlyTyrPro 59
         ||||||:::   ||||||:::||||||||||||:::::   |||:::|||
     125 CTGAAACAAACATGAATGTGAGTGAAATTATCTCTTACTGGGGATTCCCT 174

      60 CysGluGluTyrGluValAlaThrGluAspGlyTyrIleLeuSerValAs 76
            |||||||||   |||   ||||||||||||||||||||   :::||
     175 AGTGAGGAATACCTAGTTGAGACAGAAGATGGATATATTCTGTGCCTTAA 224

      76 nArgIleProArgGlyLeuValGlnProLysLysThrGlySerArgProV 93
         ||||||||||:::|||      :::   :::      |||   :::||||
     225 CCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAAACCAG 274

      93 alValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsnTrpIleSer 109
         |||||:::|||||||||||||||:::::   :::|||||||||:::::
     275 TTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGCGTCACA 324

     110 AsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAlaGlyPheAs 126
         ||||||   |||:::|||||||||||||||||||||||||||||||||||
     325 AACCTTGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGG-TTTGA 374

     126 pValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgLysHisLysT 143
         |||||||||||||||||||||||||||:::|||||||||||||||||||
     375 CGTGTGGATGGGCAACAGCAGAGGAAATACCTGGTCTCGGAAACATAAGA 424

     143 hrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMet 159
         ||||||||:::::|||||||||||||||||||||||||||||||||||||
     425 CACTCTCAGTTTCTCAGGATGAATTCTGGGCTTTCAGTTATGATGAGATG 474

     160 AlaArgPheAspLeuProAlaValIleAsnPheIleLeuGlnLysThrGl 176
         |||::::::|||||||||||||   |||||||||||||||:::|||||||
     475 GCAAAATATGACCTACCAGCTTCCATTAACTTCATTCTGAATAAAACTGG 524

     176 yGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMetGlyP 193
         |||||||:::::|||||||||||||:::|||||||||||||||:::|||||
     525 CCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTATAGGTT 574
```

```
    193 heIleAlaPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyr 209
        ||||||||||||||   :::||||||||||||:::::|||||||||:::
    575 TTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTT 624

    210 PheAlaLeuAlaProIleAlaThrValLysHisAlaLysSerProGlyTh 226
        |||||||||:::|||:::|||:::|||      :::   ||||||   ::
    625 TTTGCCCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGC 674

    226 rLysPheLeuLeuLeuProAspMetMetIleLysGlyLeuPheGlyLysL 243
        :|||:::      |||||||||   :::||||||   |||||||||   |
    675 CAAATTAGGACGATTACCAGATCATCTCATTAAGGACTTATTTGGAGACA 724

    243 ysGluPheLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTyrLeu 259
        ||||||||||   |||:::   ||||||:::   |||      ::::::
    725 AAGAATTTCTTCCCCAGAGTGCGTTTTTGAAGTGGCTGGGTACCCACGTT 774

    260 CysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMetLeuLeuLe 276
        |||   :::||||||||||   ::::::|||:::|||:::   :::|||||
    775 TGCACTCATGTCATACTGAAGGAGCTCTGTGGAAATCTCTGTTTTCTTCT 824

    276 uGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaSerValTyrA 293
        |   |||||||||   :::|||:::|||||||||||:::::::||||||:
    825 GTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGTATATA 874

    293 laAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrpSer 309
        :::::|||:::   |||||||||||||||||||:::||||||||||||
    875 CAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGC 924

    310 GlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGlySerGluTh 326
        |||||||||:::      ::::::::::|||||||||||||||||:::::
    925 CAGGCTGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGC 974

    326 rLysAsnLeuGluLysCysAsnGlnProThrProValArgTyrArgValA 343
        :||||||          ||||||     |||     |||:::|||:
    975 CAAGAATTATTTTCATTACAACCAGAGTTATCCTCCCACATACAATGTGA 1024

    343 rgAspMetThrValProThrAlaMetTrpThrGlyGlyGlnAspTrpLeu 359
        ::|||||   ||||||||||||:::|||:::||||||:::||||||||
   1025 AGGACATGCTTGTGCCGACTGCAGTCTGGAGCGGGGGTCACGACTGGCTT 1074

    360 SerAsnProGluAspValLysMetLeuLeuSerGluValThrAsnLeuIl 376
        ::::::      ||||||::::::||||||:::::::::||||||||::
   1075 GCAGATGTCTACGACGTCAATATCTTACTGACTCAGATCACCAACTTGGT 1124

    376 eTyrHisLysAsnIleProGluTrpAlaHisValAspPheIleTrpGlyL 393
        ::::|||::::::|||||||||||||   |||:::||||||||||||||||
   1125 GTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTGGGGCC 1174

    393 euAspAlaProHisArgMetTyrAsnGluIleIleHisLeuMetGlnGln 409
        ||||||||||   |||:::|||||||:::|||||||:::||||||:::::
   1175 TGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAA 1224
```

```
seq_name: gb_pr:HSU08464

seq_documentation_block:
LOCUS       HSU08464     2481 bp    mRNA            PRI       23-JUN-1994
DEFINITION  Human lysosomal acid lipase mRNA, complete cds.
ACCESSION   U08464
VERSION     U08464.1  GI:505052
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2481)
  AUTHORS   Du,H. and Gregory,G.A.
  TITLE     Structural Conservation of Putative Functional Motifs between Mouse
            and Human Lysosomal Acid Lipase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2481)
  AUTHORS   Du,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1994) Hong Du, Division of Human Gentics,
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2481
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="clone hLAL"
                     /clone_lib="HepG2 cDNA library/Lambda ZAP"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /tissue_type="liver"
     5'UTR           1. .33
     sig_peptide     34. .96
     CDS             34. .1233
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="AAB60328.1"
                     /db_xref="GI:505053"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     variation       79
                     /replace="c"
     mat_peptide     97. .1230
                     /product="lysosomal acid lipase"
     variation       100
                     /replace="g"
     3'UTR           1234. .2480
     polyA_signal    2459. .2464
     polyA_site      2480
BASE COUNT      692 a    477 c    513 g    799 t
ORIGIN


alignment_scores:
            Quality: 1298.00            Length:    400
              Ratio:  3.875              Gaps:      1
 Percent Similarity:  83.750   Percent Identity: 61.000

alignment_block:
 US-09-578-063-47 x HSU08464   ..

 Align seg 1/1  to: HSU08464  from: 1  to: 2481

     12 SerHisArgMetGluMetTrpLeuLeuIleLeuValAlaTyrMetPheGl 28
        |||   ||||||:::|||   :::|||   ||||||:::   :::
     25 AGCTCCAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCT 74

     28 nArgAsnValAsnSerValHisMetProThrLys......AlaValAspP 43
           ::::::::::|||               |||      ||||||||||
     75 CTGGACCCTGCATTCTGAGGGGTCTAGAGGGAAACTGACAGCTGTGGATC 124

     43 roGluAlaPheMetAsnIleSerGluIleIleGlnHisGlnGlyTyrPro 59
        ||||||:::   ||||||:::||||||||||||:::::   |||:::|||
    125 CTGAAACAAACATGAATGTGAGTGAAATTATCTCTTACTGGGGATTCCCT 174

     60 CysGluGluTyrGluValAlaThrGluAspGlyTyrIleLeuSerValAs 76
           |||||||||   |||   ||||||||||||||||||||||   :::||
    175 AGTGAGGAATACCTAGTTGAGACAGAAGATGGATATATTCTGTGCCTTAA 224

     76 nArgIleProArgGlyLeuValGlnProLysLysThrGlySerArgProV 93
        ||||||||||:::|||      :::   :::      |||   ::::||||
    225 CCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAAACCAG 274

     93 alValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsnTrpIleSer 109
        |||||:::||||||||||||||:::::   :::||||||||||:::::
    275 TTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACA 324
```

```
 110 AsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAlaGlyPheAs 126
     ||||||   |||:::|||||||||||||||||||||||||||||||||||
 325 AACCTTGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGA 374

 126 pValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgLysHisLysT 143
     ||||||||||||||||||||||||||:::|||||||||||||||||||||
 375 CGTGTGGATGGGCAACAGCAGAGGAAATACCTGGTCTCGGAAACATAAGA 424

 143 hrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMet 159
     ||||||||::::::|||||||||||||||||||||||||||||||||||
 425 CACTCTCAGTTTCTCAGGATGAATTCTGGGCTTTCAGTTATGATGAGATG 474

 160 AlaArgPheAspLeuProAlaValIleAsnPheIleLeuGlnLysThrGl 176
     |||:::::::|||||||||||   |||||||||||||||:::|||||||||
 475 GCAAAATATGACCTACCAGCTTCCATTAACTTCATTCTGAATAAAACTGG 524

 176 yGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMetGlyP 193
     |||||||::::::|||||||||||:::|||||||||||||||:::||||
 525 CCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTATAGGTT 574

 193 heIleAlaPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyr 209
     ||||||||||||||   :::||||||||||||:::::|||||||||:::
 575 TTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTT 624

 210 PheAlaLeuAlaProIleAlaThrValLysHisAlaLysSerProGlyTh 226
     |||||||||:::|||:::|||:::|||      :::   ||||||   ::
 625 TTTGCCCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGC 674

 226 rLysPheLeuLeuLeuProAspMetMetIleLysGlyLeuPheGlyLysL 243
     :|||:::      |||||||||   :::||||||   |||||||||   |
 675 CAAATTAGGACGATTACCAGATCATCTCATTAAGGACTTATTTGGAGACA 724

 243 ysGluPheLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTyrLeu 259
     ||||||||||||   |||:::   ||||||:::   |||      ::::::
 725 AAGAATTTCTTCCCCAGAGTGCGTTTTTGAAGTGGCTGGGTACCCACGTT 774

 260 CysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMetLeuLeuLe 276
     |||   :::|||||||||   :::::::|||:::|||:::   :::|||||
 775 TGCACTCATGTCATACTGAAGGAGCTCTGTGGAAATCTCTGTTTTCTTCT 824

 276 uGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaSerValTyrA 293
     |   |||||||||   :::|||:::|||||||||||:::::::||||||:
 825 GTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGTATATA 874

 293 laAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrpSer 309
     :::::|||:::   |||||||||||||||||||:::|||||||||||||
 875 CAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGC 924

 310 GlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGlySerGluTh 326
     |||||||||:::      ::::::::|||||||||||||||||:::::
 925 CAGGCTGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGC 974

 326 rLysAsnLeuGluLysCysAsnGlnProThrProValArgTyrArgValA 343
     :||||||            ||||||      |||      |||:::|||:
 975 CAAGAATTATTTTCATTACAACCAGAGTTATCCTCCCACATACAATGTGA 1024

 343 rgAspMetThrValProThrAlaMetTrpThrGlyGlyGlnAspTrpLeu 359
     ::|||||   ||||||||||||:::|||:::||||||:::|||||||||
1025 AGGACATGCTTGTGCCGACTGCAGTCTGGAGCGGGGGTCACGACTGGCTT 1074

 360 SerAsnProGluAspValLysMetLeuLeuSerGluValThrAsnLeuIl 376
     ::::::      ||||||:::::||||||:::::::::||||||||::
1075 GCAGATGTCTACGACGTCAATATCTTACTGACTCAGATCACCAACTTGGT 1124

 376 eTyrHisLysAsnIleProGluTrpAlaHisValAspPheIleTrpGlyL 393
     ::::|||:::::|||||||||||   |||:::||||||||||||||||||
1125 GTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTGGGGCC 1174

 393 euAspAlaProHisArgMetTyrAsnGluIleIleHisLeuMetGlnGln 409
     ||||||||||||   |||:::|||||||:::||||||:::|||||||:::::
1175 TGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAA 1224
```

```
seq_name: gb_ro:S81497

seq_documentation_block:
LOCUS       S81497       3144 bp    mRNA            ROD       02-AUG-2000
DEFINITION  lysosomal acid lipase=intracellular hydrolase [rats, Wolman, liver,
            mRNA, 3144 nt].
ACCESSION   S81497
VERSION     S81497.1  GI:1336725
KEYWORDS    .
SOURCE      Rattus sp. liver Wolman.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 3144)
  AUTHORS   Nakagawa,H., Matsubara,S., Kuriyama,M., Yoshidome,H., Fujiyama,J.,
            Yoshida,H. and Osame,M.
  TITLE     Cloning of rat lysosomal acid lipase cDNA and identification of the
            mutation in the rat model of Wolman's disease
  JOURNAL   J. Lipid Res. 36 (10), 2212-2218 (1995)
  MEDLINE   96129534
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 176569] from the original journal article.
            This sequence comes from Fig. 2.
FEATURES             Location/Qualifiers
     source          1. .3144
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
     gene            1. .3144
                     /gene="lysosomal acid lipase, LAL"
     CDS             233. .1426
                     /gene="lysosomal acid lipase, LAL"
                     /note="intracellular hydrolase; This sequence comes from
                     Fig. 2; conceptual translation presented here differs from
                     translation in publication; LAL"
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="AAB36043.2"
                     /db_xref="GI:9653291"
                     /translation="MQLLGRVICFVVGILLSGGPTGTISAVDPEANMNVTEIIMHWGY
                     PEHSVQTGDGYILGVHRIPHGRKNQFDKGPKPVVYLQWRHGFLADSSNWVTNIDNNSL
                     GFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEYWAFSFDEMAKYDLPASINYILNK
                     TGQEQLYNVGHSQGCTIGFIAFSQMPELAKKVKMFFALAPVLSLNFASGPMVKLGRLP
                     DLLLEDLFGQKQFLPQSAMVKWLSTHICTHVIMKELCANIFFLICGFNEKNLNMSRVD
                     VYTTHCPAGTSVQNMVHWTQVVKYHKLQAFDWGSSDKNYFHYNQSYPPLYSIKDMQLP
                     TALWSGGKDWLADTSDINILLTEIPTLVYHKNIPEWDHLDFIWGLDAPWRLYNEVVSL
                     MKKYQ"
BASE COUNT      852 a    690 c    754 g    848 t
ORIGIN


alignment_scores:
            Quality: 1220.50              Length:    378
              Ratio:  3.838                Gaps:      3
 Percent Similarity:  84.127   Percent Identity: 58.201

alignment_block:
 US-09-578-063-47 x S81497   ..

 Align seg 1/1  to: S81497  from: 1  to: 3144

     37 ProThr.........LysAlaValAspProGluAlaPheMetAsnIleSe 50
        ||||||         :::||||||||||||||||   ||||||:::::
    290 CCAACGGGGACCATATCAGCTGTGGACCCTGAAGCAAACATGAATGTGAC 339

     50 rGluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaT 67
        :||||||||||:::|||   |||||||||      |||::::::|||   |
    340 TGAGATAATCATGCACTGGGGATATCCA......GAGCACTCAGTACAGA 383

     67 hrGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuVal 83
        ||   ||||||||||||||:::|||:::||||||||||:::|||
    384 CAGGAGATGGCTACATTCTGGGTGTTCACCGAATCCCGCATGGGCGGAAG 433
```

```
  84 GlnProLysLysThrGlySerArgProValValLeuLeuGln......Hi 98
     :::           |||   :::|||||||||   ||||||      ||
 434 AACCAGTTTGATAAAGGTCCCAAACCAGTTGTGTATCTTCAGTGGCGACA 483

  98 sGlyLeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnS 115
     ||||:::::::::   :::|||||||||:::::::|||:::   |||||||
 484 TGGCTTCCTGGCAGATTCTAGTAACTGGGTCACAAACATTGACAACAACA 533

 115 erLeuGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsn 131
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 534 GCCTGGGCTTCATCCTGGCAGATGCTGGCTTTGATGTGTGGATGGGAAAC 583

 132 SerArgGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGl 148
     ||||||||||||:::|||||||||||||||||||||||||:::::::||
 584 AGCAGAGGAAACACCTGGTCTCGGAAGCACAAGACTCTGTCAGTTTCTCA 633

 148 nAspGluPheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeuP 165
     |||||||:::|||||||||||||:::|||||||||||||:::::|||||||
 634 GGATGAATACTGGGCCTTCAGTTTTGATGAGATGGCAAAATATGACCTCC 683

 165 roAlaValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyr 181
     |||||   ||||||:::|||||||:::|||||||||||||||:::::|||
 684 CTGCTTCAATTAACTACATTTTGAATAAAACTGGCCAAGAACAGCTGTAT 733

 182 TyrValGlyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerTh 198
        ||||||:::|||||||||   |||:::|||||||||||||||||||
 734 AATGTGGGCCACTCTCAAGGCTGCACCATAGGTTTTATAGCATTTTCACA 783

 198 rMetProGluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProI 215
      |||||||||||||||:::|||:::|||||||:::|||||||||||||||:
 784 GATGCCTGAACTGGCAAAGAAGGTTAAAATGTTTTTTGCCCTGGCTCCTG 833

 215 leAlaThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeu 231
     ::   :::::::::   |||::::::|||   :::|||:::      |||
 834 TGTTGTCTCTCAATTTTGCTTCGGGCCCGATGGTCAAATTGGGACGGTTG 883

 232 ProAspMetMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrGl 248
     ||||||:::::::::::::   |||||||||:::|||:::||||||   ||
 884 CCGGATCTGCTTCTTGAGGACTTGTTTGGGCAGAAGCAATTTCTTCCTCA 933

 248 nThrArgPheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleL 265
     |:::   :::::::::   |||      :::::::|||   :::|||||||:
 934 GAGTGCAATGGTGAAATGGCTAAGCACTCACATTTGCACGCACGTCATCA 983

 265 euAspGlnIleCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThr 281
     ::   ::::::|||:::||||||:::::::|||:::   |||||||||
 984 TGAAGGAGCTTTGCGCCAACATCTTCTTTCTGATATGTGGCTTCAATGAG 1033

 282 AsnAsnMetAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuAl 298
     :::|||:::||||||||||||:::::::||||||:::::::|||      ||
1034 AAGAATTTAAACATGTCTAGAGTGGATGTGTATACAACACACTGTCCTGC 1083

 298 aGlyThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerG 315
     ||||||||||||||||||:::::|||||:::|||:::|||:::
1084 GGGAACCTCTGTGCAAAACATGGTACACTGGACCCAGGTTGTTAAATACC 1133

 315 lyGluLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLys 331
      :::|||:::||||||||||||||||||:::   ||||||
1134 ATAAGCTTCAAGCTTTTGACTGGGGAAGCAGTGACAAGAATTATTTTCAT 1183

 332 CysAsnGlnProThrProValArgTyrArgValArgAspMetThrValPr 348
        ||||||      |||      |||   ::::::||||||   :::||
1184 TACAATCAGAGTTATCCACCCTTGTACAGCATAAAAGACATGCAACTACC 1233

 348 oThrAlaMetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAspV 365
     |||||||:::|||:::||||||:::||||||||||:::::   :::|||:
1234 CACTGCTTTATGGAGCGGGGGCAAGGACTGGCTGGCAGACACCAGTGACA 1283

 365 alLysMetLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIle 381
     :::::::::||||||:::|||:::   :::|||:::|||||||||||||||
1284 TCAATATCTTACTGACTGAAATCCCCACGTTAGTGTACCACAAGAACATT 1333

 382 ProGluTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisAr 398

      |||||||||   |||:::|||||||||||||||||||||||||||   ||
1334 CCCGAGTGGGACCATCTGGATTTTATCTGGGGTTTGGATGCCCCCTGGAG 1383

  398 gMetTyrAsnGluIleIleHisLeuMetGlnGln 409
      |:::|||||||||:::::::   ||||||::::::
1384 GCTGTATAATGAAGTAGTCAGTCTAATGAAGAAA 1417
```

```
seq_name: gb_ro:MMLYACLY

seq_documentation_block:
LOCUS       MMLYACLY     2358 bp    mRNA            ROD       24-FEB-1999
DEFINITION  M.musculus (C57 Black/6X CBA) LAL mRNA for lysosomal acid lipase.
ACCESSION   Z31689
VERSION     Z31689.1  GI:4456670
KEYWORDS    LAL; lipase; lysosomal acid lipase.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2358)
  AUTHORS   Du,H., Witte,D.P. and Grabowski,G.A.
  TITLE     Tissue and cellular specific expression of murine lysosomal acid
            lipase mRNA and protein
  JOURNAL   J. Lipid Res. 37 (5), 937-949 (1996)
  MEDLINE   96363957
REFERENCE   2  (bases 1 to 2358)
  AUTHORS   Du,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1994) Hong Du, Division of Human Genetics,
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH, 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2358
                     /organism="Mus musculus"
                     /strain="C57 Black/6X CBA"
                     /db_xref="taxon:10090"
                     /sex="Female"
                     /clone="mLAL #49"
                     /tissue_type="liver"
                     /clone_lib="Mouse liver cDNA library /Lamda-ZAP"
     gene            19. .1212
                     /gene="LAL"
     CDS             19. .1212
                     /gene="LAL"
                     /citation=[1]
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="CAA83494.1"
                     /db_xref="GI:4456671"
                     /translation="MQLQGLVFVFTIGILLSRVPTGTVSAVDPEVNMNVTEIIMRWGY
                     PGEEHSVLTGDGYILSIHRIPRGWKNHFGKGPRPVVYLQHGLLADSSNWVTNIDNSSL
                     GFLLADRGFDVWMGNSRGNTWSLKHKTLSVSQDEFWAFSFDEMAKYDLPASINYILNK
                     TGQEQIYYVGHSQGCTIGFIAFSQMPELAKKIKMFLVLAPVLSLNFASGPLLQLGRLP
                     DPLLKDMFGQKQFLPQSAMLKWLSIHVCTHVIMKELCANVFFLLCGFNEKNLNMSRVD
                     VYTTHCPAELLVQNMLHWGQVFKYRKLQAFDWGSSEKNYFHYNQSFPPSYNIKNMRLP
                     TALWSGGRDWLADINDITILLTQIPKLVYHKNIPEWDHLDFIWGLDAPWKLYDEIISL
                     MKKYQ"
     sig_peptide     19. .75
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     mat_peptide     76. .1209
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     polyA_signal    2246. .2251
     polyA_site      2358
BASE COUNT      604 a    549 c    524 g    681 t
ORIGIN


alignment_scores:
            Quality: 1206.00              Length:    415
              Ratio:  3.611                Gaps:      3
 Percent Similarity:  80.482   Percent Identity: 54.217
```

```
alignment_block:
 US-09-578-063-47 x MMLYACLY   ..    .

 Align seg 1/1  to: MMLYACLY  from: 1  to: 2358

      12 SerHisArgMetGluMetTrpLeuLeuIleLeuValAlaTyrMetPheGl 28
         :::|||      :::|||   |||      ||||||   :::::|||
          4 GCCCACTACCGGAAGATGCAACTCCAGGGCCTGGTG...TTCGTTTTCAC 50

      28 nArgAsnValAsnSerValHisMetProThr.........LysAlaValA 42
            ::::::          ::::::||||||         :::|||||||
      51 CATTGGGATCCTGCTTTCTCGGGTGCCCACGGGAACTGTATCAGCTGTGG 100

      42 spProGluAlaPheMetAsnIleSerGluIleIleGlnHisGlnGlyTyr 58
         ||||||||:::   ||||||::::::||||||||||:::::   ||||||
     101 ATCCTGAAGTCAACATGAATGTGACCGAGATAATCATGCGCTGGGGATAT 150

      59 ProCysGluGluTyrGluValAlaThrGluAspGlyTyrIleLeuSerVa 75
         |||   ||||||::::::|||   |||   ||||||||||||||||||:: 
     151 CCTGGTGAGGAACACTCGGTCCTGACAGGAGATGGCTATATTCTGAGTAT 200

      75 lAsnArgIleProArgGlyLeuValGlnProLysLysThrGlySerArgP 92
         ::::|||||||||||||||      :::            |||   ||||
     201 TCACCGAATCCCTCGTGGGTGGAAGAACCATTTTGGCAAAGGTCCCAGAC 250

      92 roValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsnTrpIle 108
         ||||||||   ||||||||||||||:::::   :::|||||||||:::
     251 CAGTTGTGTATCTTCAGCACGGCTTGCTGGCAGATTCTAGTAACTGGGTC 300

     109 SerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAlaGlyPh 125
         :::|||:::   |||:::||||||||||||:::|||||||||   |||||
     301 ACAAACATTGACAACAGCAGCTTGGGCTTCCTCCTGGCGGATCGTGGCTT 350

     125 eAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgLysHisL 142
         ||||||||||||||||||||||||||||||||:::||||||   |||||||
     351 TGATGTGTGGATGGGAAACAGCAGAGGAAACACCTGGTCTCTGAAGCACA 400

     142 ysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyrAspGlu 158
         ||||||||||:::::::||||||||||||||||||||||||:::||||||
     401 AGACTCTTTCAGTTTCTCAGGATGAGTTCTGGGCCTTCAGTTTTGATGAG 450

     159 MetAlaArgPheAspLeuProAlaValIleAsnPheIleLeuGlnLysTh 175
         ||||||:::::::||||||||||||   ||||||:::||||||:::|||||
     451 ATGGCAAAATATGACCTACCTGCTTCAATTAACTACATTTTGAATAAAAC 500

     175 rGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMetG 192
         ||||||||||:::|||||||||||||||||:::|||||||||   |||:::|
     501 TGGCCAAGAACAAATCTATTATGTGGGTCATTCTCAAGGCTGCACCATAG 550

     192 lyPheIleAlaPheSerThrMetProGluLeuAlaGlnLysIleLysMet 208
         ||||||||||||||||   ||||||||||||||:::||||||||||||||
     551 GTTTCATAGCTTTTTCACAGATGCCTGAGTTGGCAAAGAAGATTAAAATG 600

     209 TyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLysSerProGl 225
         :::::::::||||||||||:::   :::::::::   |||:::::|||
     601 TTTTTAGTCTTGGCTCCCGTGTTGTCTCTCAATTTTGCTTCAGGCCCGCT 650

     225 yThrLysPheLeuLeuLeuProAspMetMetIleLysGlyLeuPheGlyL 242
            ::::::       ||||||||   :::::::|||   :::||||||:
     651 ACTTCAATTGGGACGCTTGCCAGACCCTCTTCTCAAGGACATGTTTGGAC 700

     242 ysLysGluPheLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTyr 258
         ::|||:::||||||   |||:::   :::|||:::   |||   |||:::
     701 AGAAGCAATTTCTCCCTCAGAGTGCGATGCTGAAGTGGCTGAGCATTCAC 750

     259 LeuCysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMetLeuLe 275
         :::|||   :::|||||||:::   :::::::|||:::|||:::::::||
     751 GTTTGCACCCACGTCATCATGAAGGAGCTTTGCGCAAACGTCTTCTTCCT 800

     275 uLeuGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaSerValT 292
         ||||   |||||||||   :::|||:::||||||||||||:::::::||||
     801 GCTATGTGGCTTCAATGAGAAGAATTTAAATATGTCTAGAGTGGATGTGT 850

     292 yrAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrp 308
         ||:::::::|||      |||         |||||||||:::|||||||||
     851 ATACAACACACTGCCCTGCGGAACTTCTCGTGCAAAACATGTTACACTGG 900

     309 SerGlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGlySerGl 325
         :::|||:::   :::      :::|||:::||||||||||||||||||:: 
     901 GGTCAGGTTTTTAAATACCGTAAGCTTCAAGCTTTTGACTGGGGAAGCAG 950

     325 uThrLysAsnLeuGluLysCysAsnGlnProThrProValArgTyrArgV 342
         :   ||||||           ||||||      |||      |||::::
     951 TGAAAAGAATTATTTTCATTACAACCAGAGTTTTCCTCCCTCGTACAACA 1000

     342 alArgAspMetThrValProThrAlaMetTrpThrGlyGlyGlnAspTrp 358
         :::::::::|||   :::||||||||||:::|||:::|||||||:::||||||
    1001 TAAAAAACATGCGTCTACCCACTGCTTTGTGGAGCGGGGGCCGGGACTGG 1050

     359 LeuSerAsnProGluAspValLysMetLeuLeuSerGluValThrAsnLe 375
         |||::::::   :::|||:::   :::|||||||:::::::::   :::||
    1051 CTGGCAGACATCAATGACATCACTATCTTACTGACTCAAATCCCCAAGTT 1100

     375 uIleTyrHisLysAsnIleProGluTrpAlaHisValAspPheIleTrpG 392
         |:::|||||||||||||||||||||||||   |||:::|||||||||||||
    1101 GGTGTATCACAAGAACATTCCTGAGTGGGACCATCTGGATTTTATCTGGG 1150

     392 lyLeuAspAlaProHisArgMetTyrAsnGluIleIleHisLeuMetGln 408
         ||||||||||||   ::::::|||:::|||||||   |||||||:::
    1151 GTTTGGATGCCCCCTGGAAACTGTATGATGAAATAATCAGTCTAATGAAG 1200

     409 GlnGluGlu.................ThrAsnLeuSerGlnGly 417
         :::   :::                  |||:::|||::::::|||
    1201 AAATACCAGTGAATGCTGGACTTGAATACTTCCCTTCAGAGGGGG 1245

seq_name: gb_pat:A26689

seq_documentation_block:
LOCUS       A26689       1378 bp    DNA             PAT       05-APR-1995
DEFINITION  Precursor of rabit gastric lipase coding sequence.
ACCESSION   A26689
VERSION     A26689.1  GI:905029
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Benicourt,C., Blanchard,C. and Junien,J.L.
  TITLE     Recombinant gastric lipase from rabbit and pharmaceutical
            compositions
  JOURNAL   Patent: EP 0542629-A 9 19-MAY-1993;
            INSTITUT DE RECHERCHE JOUVEINAL
FEATURES             Location/Qualifiers
     source          1. .1378
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT      397 a    287 c    276 g    418 t
ORIGIN


alignment_scores:
            Quality: 1167.00            Length:    402
              Ratio:  3.505              Gaps:      3
 Percent Similarity:  82.836   Percent Identity: 54.478

alignment_block:
 US-09-578-063-47 x A26689   ..

 Align seg 1/1  to: A26689  from: 1  to: 1378

      16 GluMetTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAs 32
         :::|||||:::|||::::::||||||   :::::::::      :::::
      51 AAAATGTGGGTGCTTTTCATGGTGGCAGCTTTGCTATCTGCA...CTTGG 97

      32 nSerValHis.................MetProThrLysAlaValAspP 43
```

```
 98 AACTACACATGGTCTTTTTGGAAAATCAGCACCAACA.........AACC 138

139 CTGAAGTGAATATGAATATTAGTCAGATGATTTCCTACTGGGGATACCCA 188

189 AGTGAAAAATATGAAGTTGTGACTGAAGATGGTTATATCCTTGAGGTTAA 238

239 CAGAATTCCTTATGGGAAGAAAAATTCAGGGAACAGAGGCCAGAGACCCG 288

289 TCGTATTTTTGCAGCATGGTTTGCTTGCATCAGCCTCAAACTGGATCTCC 338

339 AACCTGCCCAACAACAGCCTGGCCTTTATTCTGGCAGATGCTGGTTATGG 388

389 TGTGTGGCTGGGGAACAGCAGAGGAAATACCTGGTCCAGAAGAAACTTAT 438

439 ACTATTCACCAGACTCCGTTGAATTCTGGGCTTTTAGCTTTGATGAAATG 488

489 GCTAAATATGACCTTCCAGCCACAATTGACTTCATTGTAAAGGAAACTGG 538

539 ACAGGAAAAGCTGCACTATGTTGGTCATTCCCAGGGCACCACCATTGGTT 588

589 TTATTGCCTTTTCTACAAATCCCAAGCTGGCTGAAAGAATCAAAACCTTT 638

639 TATGCATTAGCTCCAGTCGCCACTGTGAAGTATACAAAAAGCCTTGTAAA 688

689 CAAACTTAGGTTTATTCCTCCAACCATGTTCAAGATTATATTTGGTGACA 738

739 AAATATTCTACCCACACAATTTCTTTGATCAATTTCTTGCCACTCAAGTG 788

789 TGTTCCCGTGAGACACTGAATGTAATTTGCAGCAATGCCTTATTTATCAT 838

839 TTGTGGCTTTGACAGCGCAAACTTGAACATGAGTCGCTTGGATGTGTACG 888

889 TGTCACATAATCCGGCAGGAACTTCAGTTCAAAACATGCTGCACTGGACC 938

939 CAGGCTGTTAAATCTGGAAATTTTCAAGCTTTTAATTGGGGAAGTCCAGC 988

989 TCAGAATGTAGTGCACTTCAATCAGCCCACACCTCCCTACTACAATGTGA 1038
```

```
 343 rgAspMetThrValProThrAlaMetTrpThrGlyGlyGlnAspTrpLeu 359
1039 CCGCCATGAATGTGCCAATTGCAGTGTGGAGTGGTGGCAATGACTGGTTG 1088

 360 SerAsnProGluAspValLysMetLeuLeuSerGluValThrAsnLeuIl 376
1089 GCTGACCCCCAAGATGTTGACCTTTTGCTTCCAAAACTTTCTAATCTCAT 1138

 376 eTyrHisLysAsnIleProGluTrpAlaHisValAspPheIleTrpGlyL 393
1139 TTACCACAAGGAGATTCTTCCATATAATCACTTGGATTTTATCTGGGCAA 1188

 393 euAspAlaProHisArgMetTyrAsnGluIleIleHisLeuMetGlnGln 409
1189 TGAATGCTCCTCAAGAAGTTTACAATGAAATTATTTCTATGATGGCAAAA 1238

 410 GluGlu 411
1239 GATAAA 1244
```

```
seq_name: gb_pat:A26690

seq_documentation_block:
LOCUS       A26690       1378 bp    DNA             PAT       05-APR-1995
DEFINITION  Precursor of rabit gastric lipase (comp.).
ACCESSION   A26690
VERSION     A26690.1  GI:905030
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Benicourt,C., Blanchard,C. and Junien,J.L.
  TITLE     Recombinant gastric lipase from rabbit and pharmaceutical
            compositions
  JOURNAL   Patent: EP 0542629-A 10 19-MAY-1993;
            INSTITUT DE RECHERCHE JOUVEINAL
FEATURES             Location/Qualifiers
     source          1. .1378
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT      418 a    276 c    287 g    397 t
ORIGIN


alignment_scores:
            Quality: 1167.00            Length:    402
              Ratio:  3.505               Gaps:      3
 Percent Similarity:  82.836   Percent Identity: 54.478

alignment_block:
 US-09-578-063-47 x A26690/rev   ..

 Align seg 1/1  to reverse of: A26690  from: 1  to: 1378

  16 GluMetTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAs 32
1328 AAAATGTGGGTGCTTTTCATGGTGGCAGCTTTGCTATCTGCA...CTTGG 1282

  32 nSerValHis.................MetProThrLysAlaValAspP 43
1281 AACTACACATGGTCTTTTTGGAAAATCAGCACCAACA.........AACC 1241

  43 roGluAlaPheMetAsnIleSerGluIleIleGlnHisGlnGlyTyrPro 59
1240 CTGAAGTGAATATGAATATTAGTCAGATGATTTCCTACTGGGGATACCCA 1191

  60 CysGluGluTyrGluValAlaThrGluAspGlyTyrIleLeuSerValAs 76
1190 AGTGAAAAATATGAAGTTGTGACTGAAGATGGTTATATCCTTGAGGTTAA 1141
```

```
 76 nArgIleProArgGlyLeuValGlnProLysLysThrGlySerArgProV 93
    ||||||||||   |||      :::      :::   |||:::|||||||
1140 CAGAATTCCTTATGGGAAGAAAAATTCAGGGAACAGAGGCCAGAGACCCG 1091

 93 alValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsnTrpIleSer 109
    |||||:::|||||||||||||||:::::::::|||||||||||||||||||
1090 TCGTATTTTTGCAGCATGGTTTGCTTGCATCAGCCTCAAACTGGATCTCC 1041

110 AsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAlaGlyPheAs 126
    ||||||||||||||||||||:::|||||||||||||||||||||||:::
1040 AACCTGCCCAACAACAGCCTGGCCTTTATTCTGGCAGATGCTGGTTATGG 991

126 pValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgLysHisLysT 143
     ||||||:::||||||||||||||||||:::||||||||||:::::
990 TGTGTGGCTGGGGAACAGCAGAGGAAATACCTGGTCCAGAAGAAACTTAT 941

143 hrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMet 159
       |||   |||:::   ||||||||||||||||||:::||||||||
940 ACTATTCACCAGACTCCGTTGAATTCTGGGCTTTTAGCTTTGATGAAATG 891

160 AlaArgPheAspLeuProAlaValIleAsnPheIleLeuGlnLysThrGl 176
    |||:::::::||||||||||||:::|||:::||||||:::::::::|||||
890 GCTAAATATGACCTTCCAGCCACAATTGACTTCATTGTAAAGGAAACTGG 841

176 yGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMetGlyP 193
    ||||||||::::::|||||||||:::||||||||||||||||:::||||
840 ACAGGAAAAGCTGCACTATGTTGGTCATTCCCAGGGCACCACCATTGGTT 791

193 heIleAlaPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyr 209
    ||||||||||||||||   |||:::||||||:::::||||||   :::
790 TTATTGCCTTTTCTACAAATCCCAAGCTGGCTGAAAGAATCAAAACCTTT 741

210 PheAlaLeuAlaProIleAlaThrValLysHisAlaLysSerProGlyTh 226
    :::||||||||||||:::||||||||||||:::::::||||||      ::
740 TATGCATTAGCTCCAGTCGCCACTGTGAAGTATACAAAAAGCCTTGTAAA 691

226 rLysPheLeuLeuLeuProAspMetMetIleLysGlyLeuPheGlyLysL 243
    :|||:::   :::::|||      |||:::|||   :::||||||   |
690 CAAACTTAGGTTTATTCCTCCAACCATGTTCAAGATTATATTTGGTGACA 641

243 ysGluPheLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTyrLeu 259
    ||   |||      ::::::   |||   :::   |||:::      :::
640 AAATATTCTACCCACACAATTTCTTTGATCAATTTCTTGCCACTCAAGTG 591

260 CysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMetLeuLeuLe 276
    |||:::::::      |||:::   ||||||||||||   ::::::::::
590 TGTTCCCGTGAGACACTGAATGTAATTTGCAGCAATGCCTTATTTATCAT 541

276 uGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaSerValTyrA 293
    :   ||||||::::::   |||:::||||||||||||   :::||||||:
540 TTGTGGCTTTGACAGCGCAAACTTGAACATGAGTCGCTTGGATGTGTACG 491

293 laAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrpSer 309
    :::::|||:::   ||||||||||||||||||||:::||||||||||:::
490 TGTCACATAATCCGGCAGGAACTTCAGTTCAAAACATGCTGCACTGGACC 441

310 GlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGlySerGluTh 326
    |||||||||:::||||||:::::::::||||||:::|||||||||   ::
440 CAGGCTGTTAAATCTGGAAATTTTCAAGCTTTTAATTGGGGAAGTCCAGC 391

326 rLysAsnLeuGluLysCysAsnGlnProThrProValArgTyrArgValA 343
    ::::|||:::         ||||||||||||||      |||:::|||
390 TCAGAATGTAGTGCACTTCAATCAGCCCACACCTCCCTACTACAATGTGA 341

343 rgAspMetThrValProThrAlaMetTrpThrGlyGlyGlnAspTrpLeu 359
       |||:::||||||   |||:::|||:::||||||:::|||||||||
340 CCGCCATGAATGTGCCAATTGCAGTGTGGAGTGGTGGCAATGACTGGTTG 291

360 SerAsnProGluAspValLysMetLeuLeuSerGluValThrAsnLeuIl 376
    :::::::|||:::||||||   :::||||||   :::::::::||||||||
290 GCTGACCCCCAAGATGTTGACCTTTTGCTTCCAAAACTTTCTAATCTCAT 241

376 eTyrHisLysAsnIleProGluTrpAlaHisValAspPheIleTrpGlyL 393
    ||||||||||:::|||      :::   |||:::|||||||||||:::::
240 TTACCACAAGGAGATTCTTCCATATAATCACTTGGATTTTATCTGGGCAA 191

393 euAspAlaProHisArgMetTyrAsnGluIleIleHisLeuMetGlnGln 409
    :::::|||||:::::::::|||||||||||||||   :::|||   :::
190 TGAATGCTCCTCAAGAAGTTTACAATGAAATTATTTCTATGATGGCAAAA 141

410 GluGlu 411
    ::::::
140 GATAAA 135
```

```
seq_name: gb_pat:A01157

seq_documentation_block:
LOCUS       A01157       1336 bp    mRNA            PAT       29-JAN-1993
DEFINITION  R.norvegicus mRNA for prelingual lipase protein.
ACCESSION   A01157
VERSION     A01157.1  GI:14748
KEYWORDS    lipase.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1336)
  AUTHORS   Williamson,R.
  JOURNAL   Patent: WO 8500381-A 3 31-JAN-1985;
            CellTech Ltd
FEATURES             Location/Qualifiers
     source          1. .1336
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     CDS             7. .1194
                     /codon_start=1
                     /product="prelingual lipase"
                     /protein_id="CAA00136.1"
                     /db_xref="GI:14749"
                     /db_xref="SWISS-PROT:P04634"
                     /translation="MWLLLITSVISTFGGAHGLFGKLGPGNPEANMNISQMITYWGYP
                     CQEYEVVTEDGYILGVYRIPHGKNNSENIGKRPVVYLQHGLIASATNWIANLPNNSLA
                     FMLADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATINFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPTLAKKIKTFYALAPVATVKYTQSPLKKISFIPT
                     FLFKLMFGKKMFLPHTYFDDFLGTEVCSREVLDLLCSNTLFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSVQDFLHWAQLVRSGKFQAFNWGSPSQNMLHYNQKTPPEYDVSAMTVPV
                     AVWNGGNDILADPQDVAMLLPKLSNLLFHKEILAYNHLDFIWAMDAPQEVYNEMISMM
                     AED"
BASE COUNT      409 a    283 c    265 g    379 t
ORIGIN


alignment_scores:
           Quality: 1166.50            Length:    399
             Ratio:  3.556              Gaps:      1
 Percent Similarity:  82.206   Percent Identity: 54.637

alignment_block:
 US-09-578-063-47 x A01157   ..

 Align seg 1/1  to: A01157  from: 1  to: 1336

     16 GluMetTrpLeuLeuIleLeuValAlaTyrMetPheGln...ArgAsnVa 31
        :::||||||||||||:::::::::::   :::           :::::
      4 AAGATGTGGCTGTTATTAATAACAAGTGTGATATCAACATTCGGAGGTGC 53

     31 lAsnSerValHisMetProThrLysAlaValAspProGluAlaPheMetA 48
        ::::::::::                  :::|||||||||   ||||
     54 ACATGGCCTATTTGGAAAACTGGGTCCTGGAAACCCTGAAGCAAATATGA 103

     48 snIleSerGluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGlu 64
        ||||||||:::::::|||   :::   |||||||||||||:::|||||||||
    104 ATATTAGTCAGATGATAACTTACTGGGGATATCCATGTCAAGAATATGAA 153

     65 ValAlaThrGluAspGlyTyrIleLeuSerValAsnArgIleProArgGl 81
```

```
     |||:::|||||||||||||||||||||:::|||   |||||||||:::||
 154 GTTGTTACTGAAGATGGCTACATTCTGGGGGTCTACAGAATTCCTCATGG 203

  81 yLeuValGlnProLysLysThrGlySerArgProValValLeuLeuGlnH 98
     |      :::   ::::::   |||:::||||||||||||   |||||||
 204 GAAGAATAATTCTGAAAATATAGGCAAGAGACCTGTGGTGTATTTGCAGC 253

  98 isGlyLeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsn 114
     ||||||||:::::::::|||:::|||||||||:::|||||||||||||||
 254 ATGGTTTGATCGCATCAGCCACAAACTGGATTGCAAATCTACCAAACAAC 303

 115 SerLeuGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAs 131
     ||||||:::|||:::|||||||||||||||:::||||||||:::|||||
 304 AGCCTGGCCTTTATGCTAGCAGACGCTGGCTATGATGTGTGGCTGGGGAA 353

 131 nSerArgGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspG 148
     ||||||||||||:::||||||||||||:::      |||   |||:
 354 CAGTCGAGGAAATACATGGTCCAGGAAAAATGTATACTACTCACCAGACT 403

 148 lnAspGluPheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeu 164
     ::   ||||||||||||||||||:::||||||||||||:::::::||||||
 404 CAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAATATGACCTT 453

 165 ProAlaValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTy 181
     ||||||:::|||||||||||:::|||||||||||||||||||||||||::
 454 CCCGCCACAATAAACTTCATTGTACAGAAAACTGGACAAGAGAAGATACA 503

 181 rTyrValGlyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerT 198
     :|||||||||:::|||||||||||||||:::|||||||||||||||||||
 504 CTATGTTGGTCATTCTCAGGGCACCACTATTGGTTTCATTGCCTTTTCTA 553

 198 hrMetProGluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaPro 214
     ||   |||   ||||||:::||||||||||   ::::::||||||||||
 554 CCAATCCTACACTGGCCAAAAAAATCAAGACGTTTTATGCATTAGCTCCA 603

 215 IleAlaThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuLe 231
     :::|||||||||||||:::::::::|||||      |||:::   :::::
 604 GTTGCTACCGTGAAGTATACACAAAGTCCCTTGAAAAAGATTTCATTTAT 653

 231 uProAspMetMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrG 248
     :|||  ::::::::::|||   :::||||||||||||   ||||||   :
 654 TCCTACATTTCTTTTCAAGCTTATGTTTGGCAAGAAAATGTTCCTGCCCC 703

 248 lnThrArgPheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIle 264
     ::|||   |||         |||         :::|||::::::   :::
 704 ACACCTACTTTGATGACTTTCTTGGTACCGAAGTGTGCTCACGGGAGGTT 753

 265 LeuAspGlnIleCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnTh 281
     ||||||   :::||||||||   ::::::::::::   ||||||:::
 754 CTAGATCTTCTCTGCAGCAACACTTTATTCATCTTCTGTGGATTTGACAA 803

 281 rAsnAsnMetAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuA 298
      :::|||:::|||:::||||||   :::||||||   :::|||:::   |
 804 GAAAAACTTAAATGTGAGTCGTTTTGATGTGTATCTAGGGCATAATCCAG 853

 298 laGlyThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSer 314
     ||||||||||||||||:::::|||||||||:::|||   |||:::|||
 854 CAGGGACATCTGTTCAAGACTTTCTCCACTGGGCACAGCTTGTTAGATCT 903

 315 GlyGluLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLy 331
     |||:::::::::|||||:::||||||||   :::::::|||:::
 904 GGGAAATTTCAAGCCTTCAACTGGGGAAGCCCATCCCAGAACATGTTACA 953

 331 sCysAsnGlnProThrProValArgTyrArgValArgAspMetThrValP 348
       ||||||   ||||||   :::|||   |||      ||||||||||
 954 CTACAACCAGAAAACGCCTCCTGAATATGATGTGTCAGCCATGACTGTGC 1003

 348 roThrAlaMetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAsp 364
     ||:::|||:::|||:::||||||:::|||   |||:::::|||:::|||
1004 CAGTTGCAGTGTGGAACGGTGGCAATGACATCCTGGCTGATCCCCAAGAT 1053

 365 ValLysMetLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIl 381
     |||   |||||||||   :::::::::||||||:::::::||||||:::||
```

```
1054 GTCGCCATGCTGCTTCCCAAACTCTCCAACCTCCTGTTCCATAAGGAGAT 1103

 381 eProGluTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisA 398
     |      :::   |||:::|||||||||||||:::::|||||||||:::
1104 TCTTGCCTACAATCACCTGGACTTCATCTGGGCAATGGATGCCCCTCAAG 1153

 398 rgMetTyrAsnGluIleIleHisLeuMetGlnGlnGluGluThrAsn 413
     :::::|||||||||:::|||  :::|||   ::::::      |||
1154 AGGTTTACAATGAGATGATTTCCATGATGGCAGAAGACTAAAAGAAT 1200
```

```
seq_name: gb_ro:RNLIP

seq_documentation_block:
LOCUS       RNLIP          1355 bp    mRNA            ROD       30-MAR-1995
DEFINITION  Rat mRNA for lingual lipase.
ACCESSION   X02309
VERSION     X02309.1  GI:56595
KEYWORDS    glycoprotein; lipase.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1355)
  AUTHORS   Docherty,A.J., Bodmer,M.W., Angal,S., Verger,R., Riviere,C.,
            Lowe,P.A., Lyons,A., Emtage,J.S. and Harris,T.J.
  TITLE     Molecular cloning and nucleotide sequence of rat lingual lipase
            cDNA
  JOURNAL   Nucleic Acids Res. 13 (6), 1891-1903 (1985)
  MEDLINE   85215587
COMMENT     Data kindly reviewed (12-FEB-1986) by A.J.P. Docherty.
FEATURES             Location/Qualifiers
     source          1. .1355
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     sig_peptide     49. .102
                     /product="put. signal peptide (aa -18 to -1)"
     CDS             49. .1236
                     /note="lipase precursor"
                     /codon_start=1
                     /protein_id="CAA26179.1"
                     /db_xref="GI:56596"
                     /db_xref="SWISS-PROT:P04634"
                     /translation="MWLLLITSVISTFGGAHGLFGKLGPGNPEANMNISQMITYWGYP
                     CQEYEVVTEDGYILGVYRIPHGKNNSENIGKRPVVYLQHGLIASATNWIANLPNNSLA
                     FMLADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATINFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPTLAKKIKTFYALAPVATVKYTQSPLKKISFIPT
                     FLFKLMFGKKMFLPHTYFDDFLGTEVCSREVLDLLCSNTLFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSVQDFLHWAQLVRSGKFQAFNWGSPSQNMLHYNQKTPPEYDVSAMTVPV
                     AVWNGGNDILADPQDVAMLLPKLSNLLFHKEILAYNHLDFIWAMDAPQEVYNEMISMM
                     AED"
     mat_peptide     103. .1233
                     /product="lipase (aa 1-377)"
     misc_feature    505. .507
                     /note="aa 153 pot. lipid binding site"
     misc_feature    1355
                     /note="polyA site"
BASE COUNT      400 a    291 c    276 g    387 t      1 others
ORIGIN


alignment_scores:
            Quality: 1166.50          Length:    399
              Ratio:  3.556            Gaps:      1
 Percent Similarity:  82.206   Percent Identity: 54.637

alignment_block:
 US-09-578-063-47 x RNLIP   ..

 Align seg 1/1  to: RNLIP  from: 1  to: 1355

      16 GluMetTrpLeuLeuIleLeuValAlaTyrMetPheGln...ArgAsnVa 31
         :::||||||||||||:::::::::::   :::           :::::
```


```
seq_name: gb_om:CFLIPASE

seq_documentation_block:
LOCUS       CFLIPASE     1651 bp    mRNA            MAM       18-JUN-1998
DEFINITION  Canis familiaris mRNA for lipase.
ACCESSION   Y13899
VERSION     Y13899.1  GI:2204112
KEYWORDS    DGL gene; gastric lipase; triacylglycerol lipase.
SOURCE      dog.
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 1651)
  AUTHORS   Vaganay,S., Joliff,G., Bertaux,O., Toselli,E., Devignes,M.D. and
            Benicourt,C.
  TITLE     The complete cDNA sequence encoding dog gastric lipase
  JOURNAL   DNA Seq. 8 (4), 257-262 (1998)
  MEDLINE   99450174
REFERENCE   2  (bases 1 to 1651)
  AUTHORS   Benicourt,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1997) C. Benicourt, Ecole Normale Superieure de
            Cachan, L.I.R.B.A (Laboratoire Interdisciplinaire de Recherche en
            Biologie Appliquee), 61 avenue du President Wilson, F-94235 Cachan
            Cedex, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1651
                     /organism="Canis familiaris"
                     /db_xref="taxon:9615"
                     /dev_stage="adult"
                     /tissue_type="stomach mucosa"
     gene            54. .1250
                     /gene="DGL"
     CDS             54. .1250
                     /gene="DGL"
                     /EC_number="3.1.1.3"
                     /codon_start=1
                     /product="triacylglycerol lipase"
                     /protein_id="CAA74198.1"
                     /db_xref="GI:2204113"
                     /db_xref="SWISS-PROT:P80035"
                     /translation="MWLLLTAASVISTLGTTHGLFGKLHPTNPEVTMNISQMITYWGY
                     PAEEYEVVTEDGYILGIDRIPYGRKNSENIGRRPVAFLQHGLLASATNWISNLPNNSL
                     AFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFILKK
                     TGQDKLHYVGHSQGTTIGFIAFSTNPKLAKRIKTFYALAPVATVKYTETLLNKLMLVP
                     SFLFKLIFGNKIFYPHHFFDQFLATEVCSRETVDLLCSNALFIICGFDTMNLNMSRLD
                     VYLSHNPAGTSVQNVLHWSQAVKSGKFQAFDWGSPVQNMMHYHQSMPPYYNLTDMHVP
                     IAVWNGGNDLLADPHDVDLLLSKLPNLIYHRKIPPYNHLDFIWAMDAPQAVYNEIVSM
                     MGTDNK"
     sig_peptide     54. .110
                     /gene="DGL"
                     /evidence=experimental
```

```
     mat_peptide     111. .1247
                     /gene="DGL"
     polyA_signal    1602. .1607
BASE COUNT      432 a    419 c    352 g    448 t
ORIGIN


alignment_scores:
            Quality: 1154.00             Length:    398
              Ratio:  3.486               Gaps:      5
 Percent Similarity:  83.166   Percent Identity: 54.271

alignment_block:
 US-09-578-063-47 x CFLIPASE   ..

 Align seg 1/1  to: CFLIPASE  from: 1  to: 1651

      17 MetTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAsnSe 33
         ||||||||||||:::   :::|||   :::::::::   ::::::::::
      54 ATGTGGCTGCTATTAACAGCGGCGAGTGTGATATCT...ACACTTGGAAC 100

      33 rValHis.................MetProThrLysAlaValAspProG 44
         ::::|||                 ||||||        :::|||||
     101 CACACATGGTTTATTTGGAAAATTACATCCCACA.........AACCCTG 141

      44 luAlaPheMetAsnIleSerGluIleIleGlnHisGlnGlyTyrProCys 60
         ||:::   ||||||||||||:::::|||   :::   |||||||||:::
     142 AAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCT 191

      61 GluGluTyrGluValAlaThrGluAspGlyTyrIleLeuSerValAsnAr 77
         ||||||||||||||:::|||||||||||||||||||||||:::::::::||
     192 GAGGAATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAG 241

      77 gIleProArgGlyLeuValGlnProLysLysThrGlySerArgProValV 94
         |||||||   |||      :::   ::::::   |||   |||||||||:
     242 AATTCCTTATGGGAGGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTG 291

      94 alLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsnTrpIleSerAsn 110
         :::::|||||||||||||||||:::::::::|||:::|||||||||||||||
     292 CATTTTTGCAACACGGTTTGCTCGCATCAGCCACAAACTGGATCTCCAAC 341

     111 LeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAlaGlyPheAspVa 127
         ||||||||||||||||||:::|||||||||||||||||||||:::||||||
     342 CTGCCCAACAACAGCCTGGCCTTCATCCTGGCCGACGCCGGGTACGACGT 391

     127 lTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgLysHisLysThrL 144
         ||||:::|||||||||||||||||||:::|||:::|||::::::
     392 GTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATCTGTACT 441

     144 euSerIleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMetAla 160
           |||   |||:::   ||||||||||||||||||:::||||||||||||
     442 ACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCT 491

     161 ArgPheAspLeuProAlaValIleAsnPheIleLeuGlnLysThrGlyGl 177
         ::::::||||||||||||:::|||:::||||||||:::||||||||||||
     492 AAATATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACA 541

     177 nGluLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMetGlyPheI 194
         |:::|||::::::|||||||||:::||||||||||||||||:::|||||||
     542 GGACAAGCTACACTACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCA 591

     194 leAlaPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyrPhe 210
         ||||||||||||   |||:::||||||:::::||||||   ::::::
     592 TCGCCTTTTCCACCAATCCCAAGCTGGCGAAACGGATCAAAACCTTCTAT 641

     211 AlaLeuAlaProIleAlaThrValLysHisAlaLysSerProGlyThrLy 227
         ||||||||||||:::||||||||||||:::::::::::      :::||
     642 GCATTAGCTCCCGTTGCCACCGTGAAGTACACCGAAACCCTGTTAAACAA 691

     227 sPheLeuLeuLeuProAspMetMetIleLysGlyLeuPheGlyLysLysG 244
         |::::::|||:::|||:::::::::::::|||   :::||||||:::|||
     692 ACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAAACAAAA 741

     244 luPheLeuTyrGlnThrArgPheLeuArgGln...LeuValIleTyrLeu 259
           |||   |||       :::|||:::   |||   |||:::      :::
     742 TATTC...TACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTA 788

     260 CysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMetLeuLeuLe 276
         |||::::::      :::|||   :::|||||||||   ::::::::::
     789 TGCTCCCGCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCAT 838

     276 uGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaSerValTyrA 293
         :   ||||||:::|||   |||:::||||||||||||   :::||||||
     839 TTGTGGATTTGACACTATGAACTTGAACATGAGTCGCTTGGATGTGTATC 888

     293 laAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrpSer 309
         :::|||:::   |||||||||||||||||||:::|||||||||||||
     889 TGTCACATAATCCAGCAGGAACATCGGTTCAGAACGTGCTCCACTGGTCC 938

     310 GlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGlySerGluTh 326
         |||||||||:::||||||:::::::::||||||||||||||||||   ::
     939 CAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTGACTGGGGAAGCCCAGT 988

     326 rLysAsnLeuGluLysCysAsnGlnProThrProValArgTyrArgValA 343
         ::::|||:::          :::|||      |||      |||::::::
     989 TCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACCTGA 1038

     343 rgAspMetThrValProThrAlaMetTrpThrGlyGlyGlnAspTrpLeu 359
           ||||||   ||||||   |||:::|||:::||||||:::|||   |||
    1039 CAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTG 1088

     360 SerAsnProGluAspValLysMetLeuLeuSerGluValThrAsnLeuIl 376
         :::::::|||:::|||||||   :::|||||||||:::::   ||||||||
    1089 GCCGACCCTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCAT 1138

     376 eTyrHisLysAsnIleProGluTrpAlaHisValAspPheIleTrpGlyL 393
         |||||||:::::|||||||   :::   |||:::||||||||||||:::
    1139 TTACCACAGGAAGATTCCTCCTTACAATCACTTGGACTTTATCTGGGCCA 1188

     393 euAspAlaProHisArgMetTyrAsnGluIleIleHisLeuMet 407
         ::||||||||||::   :::||||||||||||||:::   :::|||
    1189 TGGATGCCCCTCAAGCGGTTTACAATGAAATTGTTTCCATGATG 1232

seq_name: gb_pat:AR039022

seq_documentation_block:
LOCUS       AR039022     1137 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 2 from patent US 5807726.
ACCESSION   AR039022
VERSION     AR039022.1  GI:5958385
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1137)
  AUTHORS   Blanchard,C., Benicourt,C. and Junien,J.
  TITLE     Nucleic acids encoding dog gastric lipase and their use for the
            production of polypeptides
  JOURNAL   Patent: US 5807726-A 2 15-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .1137
                     /organism="unknown"
BASE COUNT      296 a    308 c    249 g    284 t
ORIGIN


alignment_scores:
            Quality: 1141.50             Length:    372
              Ratio:  3.624               Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
 US-09-578-063-47 x AR039022   ..

 Align seg 1/1  to: AR039022  from: 1  to: 1137
```

```
 37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
 19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

 53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
 60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

 70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

 87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 359

153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409

170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352

907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

402 uIleIleHisLeuMet 407
1107 AATTGTTTCCATGATG 1122

seq_name: gb_pat:AR092633

seq_documentation_block:
LOCUS       AR092633     1137 bp    DNA             PAT       08-SEP-2000
DEFINITION  Sequence 2 from patent US 5998189.
ACCESSION   AR092633
VERSION     AR092633.1  GI:10019385
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1137)
  AUTHORS   Blanchard,C., Benicourt,C. and Junien,J.
  TITLE     Polypeptide derivatives of dog gastric lipase and pharmaceutical
            compositions containing same
  JOURNAL   Patent: US 5998189-A 2 07-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .1137
                     /organism="unknown"
BASE COUNT      296 a    308 c    249 g    284 t
ORIGIN


alignment_scores:
            Quality: 1141.50            Length:    372
              Ratio:  3.624               Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
 US-09-578-063-47 x AR092633   ..

 Align seg 1/1  to: AR092633  from: 1  to: 1137

      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
      19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
      60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

      70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
     110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

      87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
     160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

     103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
     210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

     120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
```

```
260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 359

153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409

170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

402 uIleIleHisLeuMet 407
1107 AATTGTTTCCATGATG 1122

seq_name: gb_pat:AR039023
```

```
seq_documentation_block:
LOCUS       AR039023     1140 bp    DNA              PAT       29-SEP-1999
DEFINITION  Sequence 4 from patent US 5807726.
ACCESSION   AR039023
VERSION     AR039023.1  GI:5958386
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1140)
  AUTHORS   Blanchard,C., Benicourt,C. and Junien,J.
  TITLE     Nucleic acids encoding dog gastric lipase and their use for the
            production of polypeptides
  JOURNAL   Patent: US 5807726-A 4 15-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .1140
                     /organism="unknown"
BASE COUNT      297 a    308 c    250 g    285 t
ORIGIN


alignment_scores:
            Quality: 1141.50              Length:    372
              Ratio:  3.624                Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
 US-09-578-063-47 x AR039023   ..

 Align seg 1/1  to: AR039023  from: 1  to: 1140

 37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
 22 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 62

 53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
 63 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 112

 70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
113 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 162

 87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
163 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 212

103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
213 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 262

120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
263 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 312

137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
313 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 362

153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
363 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 412

170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
413 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 462

187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
463 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 512

203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
```

```
 513 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 562

 220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
     ::::::::::::      :::|||:::::::|||:::|||:::::::::::::
 563 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 612

 237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
     |||   :::||||||:::|||   |||   |||      :::|||:::
 613 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 659

 253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
      |||   |||:::      :::|||::::::      :::|||   :::|
 660 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 709

 269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
     ||||||||   ::::::::::::   ||||||:::|||   |||:::|||
 710 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 759

 286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
     |||||||||   :::||||||   :::|||:::   ||||||||||||||
 760 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 809

 302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
     |||||||:::||||||||||||||||||||||:::||||||:::::::::|
 810 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 859

 319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
     ||||||||||||||||   ::::::|||:::         :::|||
 860 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 909

 336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
        |||      |||::::::   ||||||   ||||||   |||:::||
 910 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 959

 352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
     |:::||||||:::|||   |||::::::|||:::||||||   :::||||
 960 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1009

 369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
     |||||:::::   ||||||||||||||:::::::||||||   :::
1010 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1059

 386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
     |||:::||||||||||||:::::::||||||||:::   :::||||||||
1060 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1109

 402 uIleIleHisLeuMet 407
     ||||:::   :::|||
1110 AATTGTTTCCATGATG 1125
```

OM of: US-09-578-063-47 to: N_Geneseq_1101:* out_format : pfs

Date: Jan 31, 2002 12:08 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09578063/runat_30012002_092151_26344/app_query.fasta_1.488
-DB=N_Geneseq_1101 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09578063_@CGN1_1_308 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-578-063-47
Query length: 423
Database: N_Geneseq_1101:*
Database sequences: 930621
Database length: 428662619
Search time (sec): 171.190000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF45132	+	2247.00	4488.65	6.1e-242	1269	!	..
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF45131	+	2247.00	4483.08	1.2e-241	2044	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH02906	+	1306.00	2588.07	4.4e-136	2493	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ42310	+	1167.00	2315.42	6.8e-121	1378	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:AAN50385	+	1158.50	2298.69	5.8e-120	1336	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ68388	+	1141.50	2262.93	5.7e-118	1528	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT58915	+	1141.50	2262.93	5.7e-118	1528	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ68389	+	1141.50	2262.90	5.7e-118	1531	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:AAN60685	+	1124.00	2229.03	4.4e-116	1367	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:AAN60566	+	1124.00	2229.03	4.4e-116	1367	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT58916	+	1124.00	2229.03	4.4e-116	1367	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH57420	+	1119.00	2216.46	2.2e-115	1695	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF28689	+	695.00	1366.67	4.8e-68	1308	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF28679	+	691.00	1365.95	5.2e-68	699	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA44349	+	596.00	1176.33	1.9e-57	617	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ34958	+	514.50	1000.44	1.2e-47	1718	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ34956	+	502.00	977.01	2.4e-46	1483	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA43744	+	473.50	930.82	9.0e-44	572	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ34959	+	465.00	902.95	3.2e-42	1438	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ34953	+	398.00	774.21	4.8e-35	859	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ34952	+	309.50	598.52	2.9e-25	705	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH57212	-	272.00	529.66	2.0e-21	402	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ60606	+	271.50	473.86	2.6e-18	43676		
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC93827	+	230.50	442.55	1.4e-16	549	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF12152	+	173.50	324.94	5.0e-10	707	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ34954	+	160.50	302.63	8.8e-09	509	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV35431	+	155.00	303.85	7.5e-09	178	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ34950	+	150.00	276.97	2.4e-07	751	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ34955	+	138.00	264.66	1.2e-06	273	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ34951	+	138.00	254.58	4.2e-06	647	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT28743	+	117.00	208.13	0.0016	927	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT73350	+	117.00	198.98	0.0052	2028	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT28741	+	116.50	207.99	0.0016	861	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT73347	+	116.50	207.99	0.0016	861	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV74384	+	110.00	173.36	0.1401	5445	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT73289	+	105.00	177.07	0.0870	1676	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT62362	+	105.00	177.07	0.0870	1676	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79322	+	102.00	176.61	0.0924	1041	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI58295	+	100.00	165.66	0.3760	1882	!	
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ09252	-	99.50	131.52	29.97	32042		
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF30011	-	99.50	131.52	29.97	3204		
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH67490	+	99.00	162.74	0.5467	2034		
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH98291	+	99.00	142.48	7.36	1152		
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH68531	-	99.00	102.57	1.2e+03	3499		
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH68532	-	99.00	102.57	1.2e+03	3499		

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF45132

seq_documentation_block:
ID AAF45132 standard; cDNA; 1269 BP.
XX
AC AAF45132;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294 ORF.
XX
KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety; ss.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
DR P-PSDB; AAB66065.
XX
PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease -
XX
PS Claim 1; Fig 6; 359pp; English.
XX
CC The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other
CC neurological and cerebrovascular disorders. The CNS disorders include
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC autonomic function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder.
XX
SQ Sequence 1269 BP; 358 A; 268 C; 309 G; 334 T; 0 other;

alignment_scores:
Quality: 2247.00 Length: 423
Ratio: 5.312 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-578-063-47 x AAF45132 ..

Align seg 1/1 to: AAF45132 from: 1 to: 1269

```
  1 MetLeuGluThrLeuSerArgGlnTrpIleValSerHisArgMetGluMe 17
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 ATGTTGGAAACCTTGTCAAGACAGTGGATTGTCTCACACAGAATGGAAAT 50

 17 tTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAsnSerV 34
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 GTGGCTTCTGATTCTGGTGGCGTATATGTTCCAGAGAAATGTGAATTCAG 100

 34 alHisMetProThrLysAlaValAspProGluAlaPheMetAsnIleSer 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 TACATATGCCAACTAAAGCTGTGGACCCAGAAGCATTCATGAATATTAGT 150

 51 GluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaTh 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 GAAATCATCCAACATCAAGGCTATCCCTGTGAGGAATATGAAGTCGCAAC 200

 67 rGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuValG 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 TGAAGATGGGTATATCCTTTCTGTTAACAGGATTCCTCGAGGCCTAGTGC 250

 84 lnProLysLysThrGlySerArgProValValLeuLeuGlnHisGlyLeu 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 AACCTAAGAAGACAGGTTCCAGGCCTGTGGTGTTACTGCAGCATGGCCTA 300

101 ValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGl 117
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 GTTGGAGGTGCTAGCAACTGGATTTCCAACCTGCCCAACAATAGCCTGGG 350

117 yPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgG 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 CTTCATTCTGGCAGATGCTGGTTTTGACGTGTGGATGGGGAACAGCAGGG 400

134 lyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAspGlu 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
401 GAAACGCCTGGTCTCGAAAACACAAGACACTCTCCATAGACCAAGATGAG 450

151 PheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeuProAlaVa 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 TTCTGGGCTTTCAGTTATGATGAGATGGCTAGGTTTGACCTTCCTGCAGT 500

167 lIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValG 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
501 GATAAACTTTATTTTGCAGAAAACGGGCCAGGAAAAGATCTATTATGTCG 550

184 lyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMetPro 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
551 GCTATTCACAGGGCACCACCATGGGCTTTATTGCATTTTCCACCATGCCA 600

201 GluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaTh 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||
601 GAGCTGGCTCAGAAAATCAAAATGTATTTTGCTTTAGCACCCATAGCCAC 650

217 rValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeuProAspM 234
    ||||||||||||||||||||||||||||||||||||||||||||||||||
651 TGTTAAGCATGCAAAAAGCCCCGGGACCAAATTTTTGTTGCTGCCAGATA 700

234 etMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArg 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
701 TGATGATCAAGGGATTGTTTGGCAAAAAAGAATTTCTGTATCAGACCAGA 750

251 PheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGl 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||
751 TTTCTCAGACAACTTGTTATTTACCTTTGTGGCCAGGTGATTCTTGATCA 800

267 nIleCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnM 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||
801 GATTTGTAGTAATATCATGTTACTTCTGGGTGGATTCAACACCAACAATA 850

284 etAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThr 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
851 TGAACATGAGCCGAGCAAGTGTATATGCTGCCCACACTCTTGCTGGAACA 900

301 SerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLe 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||
901 TCTGTGCAAAATATTCTACACTGGAGCCAGGCAGTGAATTCTGGTGAACT 950

317 uArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnG 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||
951 CCGGGCATTTGACTGGGGGAGTGAGACCAAAAATCTGGAAAAATGCAATC 1000

334 lnProThrProValArgTyrArgValArgAspMetThrValProThrAla 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 AGCCAACTCCTGTAAGGTACAGAGTCAGAGATATGACGGTCCCTACAGCA 1050

351 MetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMe 367
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 ATGTGGACAGGAGGTCAGGACTGGCTTTCAAATCCAGAAGACGTGAAAAT 1100

367 tLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluT 384
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1101 GCTGCTCTCTGAGGTGACCAACCTCATCTACCATAAGAATATTCCTGAAT 1150

384 rpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyr 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1151 GGGCTCACGTGGATTTCATCTGGGGTTTGGATGCTCCTCACCGTATGTAC 1200

401 AsnGluIleIleHisLeuMetGlnGlnGluGluThrAsnLeuSerGlnGl 417
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1201 AATGAAATCATCCATCTGATGCAGCAGGAGGAGACCAACCTTTCCCAGGG 1250

417 yArgCysGluAlaValLeu 423
    |||||||||||||||||||
1251 ACGGTGTGAGGCCGTATTG 1269
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF45131

seq_documentation_block:

ID AAF45131 standard; cDNA; 2044 BP.
XX
AC AAF45131;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294 cDNA.
XX
KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety; ss.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
DR P-PSDB; AAB66065.
XX
PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease -
XX
PS Claim 1; Fig 6; 359pp; English.

```
XX
CC   The present invention relates to TANGO or INTERCEPT proteins and coding
CC   sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC   AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC   sequences are useful for the treatment of neurological disorders such as
CC   central nervous system (CNS) disorders, CNS-related disorders, focal
CC   brain disorders, global-diffuse cerebral disorders and other
CC   neurological and cerebrovascular disorders. The CNS disorders include
CC   Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC   lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC   autonomic function disorders such as hypertension and sleep disorders;
CC   neuropsychiatric disorders, psychoactive substance use disorders,
CC   anxiety, and bipolar affective disorder.
XX
SQ   Sequence 2044 BP; 618 A; 401 C; 460 G; 565 T; 0 other;


alignment_scores:
           Quality: 2247.00            Length:    423
             Ratio:  5.312              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-578-063-47 x AAF45131   ..

 Align seq 1/1  to: AAF45131  from: 1  to: 2044

       1 MetLeuGluThrLeuSerArgGlnTrpIleValSerHisArgMetGluMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     126 ATGTTGGAAACCTTGTCAAGACAGTGGATTGTCTCACACAGAATGGAAAT 175

      17 tTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAsnSerV 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     176 GTGGCTTCTGATTCTGGTGGCGTATATGTTCCAGAGAAATGTGAATTCAG 225

      34 alHisMetProThrLysAlaValAspProGluAlaPheMetAsnIleSer 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     226 TACATATGCCAACTAAAGCTGTGGACCCAGAAGCATTCATGAATATTAGT 275

      51 GluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     276 GAAATCATCCAACATCAAGGCTATCCCTGTGAGGAATATGAAGTCGCAAC 325

      67 rGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuValG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     326 TGAAGATGGGTATATCCTTTCTGTTAACAGGATTCCTCGAGGCCTAGTGC 375

      84 lnProLysLysThrGlySerArgProValValLeuGluGlnHisGlyLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     376 AACCTAAGAAGACAGGTTCCAGGCCTGTGGTGTTACTGCAGCATGGCCTA 425

     101 ValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     426 GTTGGAGGTGCTAGCAACTGGATTTCCAACCTGCCCAACAATAGCCTGGG 475

     117 yPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     476 CTTCATTCTGGCAGATGCTGGTTTTGACGTGTGGATGGGGAACAGCAGGG 525

     134 lyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAspGlu 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     526 GAAACGCCTGGTCTCGAAAACACAAGACACTCTCCATAGACCAAGATGAG 575

     151 PheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeuProAlaVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     576 TTCTGGGCTTTCAGTTATGATGAGATGGCTAGGTTTGACCTTCCTGCAGT 625

     167 lIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     626 GATAAACTTTATTTTGCAGAAAACGGGCCAGGAAAAGATCTATTATGTCG 675

     184 lyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMetPro 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     676 GCTATTCACAGGGCACCACCATGGGCTTTATTGCATTTTCCACCATGCCA 725

     201 GluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     726 GAGCTGGCTCAGAAAATCAAAATGTATTTTGCTTTAGCACCCATAGCCAC 775

     217 rValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeuProAspM 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     776 TGTTAAGCATGCAAAAAGCCCCGGGACCAAATTTTTGTTGCTGCCAGATA 825

     234 etMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArg 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     826 TGATGATCAAGGGATTGTTTGGCAAAAAAGAATTTCTGTATCAGACCAGA 875

     251 PheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     876 TTTCTCAGACAACTTGTTATTTACCTTTGTGGCCAGGTGATTCTTGATCA 925

     267 nIleCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnM 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     926 GATTTGTAGTAATATCATGTTACTTCTGGGTGGATTCAACACCAACAATA 975

     284 etAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThr 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     976 TGAACATGAGCCGAGCAAGTGTATATGCTGCCCACACTCTTGCTGGAACA 1025

     301 SerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1026 TCTGTGCAAAATATTCTACACTGGAGCCAGGCAGTGAATTCTGGTGAACT 1075

     317 uArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1076 CCGGGCATTTGACTGGGGGAGTGAGACCAAAAATCTGGAAAAATGCAATC 1125

     334 lnProThrProValArgTyrArgValArgAspMetThrValProThrAla 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1126 AGCCAACTCCTGTAAGGTACAGAGTCAGAGATATGACGGTCCCTACAGCA 1175

     351 MetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMe 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1176 ATGTGGACAGGAGGTCAGGACTGGCTTTCAAATCCAGAAGACGTGAAAAT 1225

     367 tLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluT 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1226 GCTGCTCTCTGAGGTGACCAACCTCATCTACCATAAGAATATTCCTGAAT 1275

     384 rpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyr 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1276 GGGCTCACGTGGATTTCATCTGGGGTTTGGATGCTCCTCACCGTATGTAC 1325

     401 AsnGluIleIleHisLeuMetGlnGlnGluGluThrAsnLeuSerGlnGl 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1326 AATGAAATCATCCATCTGATGCAGCAGGAGGAGACCAACCTTTCCCAGGG 1375

     417 yArgCysGluAlaValLeu 423
         |||||||||||||||||||
    1376 ACGGTGTGAGGCCGTATTG 1394

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH02906

seq_documentation_block:
ID   AAH02906 standard; DNA; 2493 BP.
XX
AC   AAH02906;
XX
DT   15-JUN-2001  (first entry)
XX
DE   Human shear stress-response coding sequence SEQ ID NO: 65.
XX
KW   Human; shear stress-response protein; vascular disease;
KW   arteriosclerosis; ds.
XX
OS   Homo sapiens.
```

```
XX
PN   WO200125427-A1.
XX
PD   12-APR-2001.
XX
PF   02-OCT-2000; 2000WO-JP06840.
XX
PR   01-OCT-1999;   99JP-0280976.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (NOJI/) NOJIMA H.
XX
PI   Nojima H,  Yoshisue H,  Obayashi M,  Ota T,  Kawabata A,  Sakurada K;
PI   Kuga T,  Sekine S,  Nakamura Y,  Sugano S;
XX
DR   WPI; 2001-266308/27.
DR   P-PSDB; AAB90783.
XX
PT   DNA sequences, proteins encoded by them and antibodies against them
PT   useful in diagnosis and treatment of vascular disease caused by
PT   arteriosclerosis -
XX
PS   Claim 20; Page 398-402; 678pp; Japanese.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human shear stress response proteins. These are useful in the
CC   diagnosis, treatment and screening of vascular diseases caused by
CC   arteriosclerosis, including heart failure, post-PTCA restenosis and
CC   hypertension.
XX
SQ   Sequence 2493 BP; 692 A; 484 C; 516 G; 801 T; 0 other;


alignment_scores:
            Quality: 1306.00              Length:    410
              Ratio:  3.841                Gaps:      2
 Percent Similarity:  82.927    Percent Identity: 60.732

alignment_block:
 US-09-578-063-47 x AAH02906   ..

 Align seq 1/1  to: AAH02906  from: 1  to: 2493

       2 LeuGluThrLeuSerArgGlnTrpIleValSerHisArgMetGluMetTr 18
         |||||||||   :::||||||          |||   ||||||:::|||
       8 CTCGAGACAGCGGCCCGGCAG......GACAGCTCCAGAATGAAAATGCG 51

      18 pLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAsnSerValH 35
         :::|||   |||||||:::   :::          :::::::|||
      52 GTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGCCCCTGCATTCTGAGG 101

      35 isMetProThrLys......AlaValAspProGluAlaPheMetAsnIle 49
                    |||      ||||||||||||||||:::   ||||||:::
     102 GGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTG 151

      50 SerGluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAl 66
         ||||||||||||:::::   |||:::|||   |||||||||   |||
     152 AGTGAAATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGA 201

      66 aThrGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuV 83
         ||||||||||||||||||||   :::||||||||||||:::|||
     202 GACAGAAGATGGATATATTCTGTGCCTTAACCGAATTCCTCATGGGAGGA 251

      83 alGlnProLysLysThrGlySerArgProValValLeuLeuGlnHisGly 99
         :::   :::      |||   :::||||||||:::||||||||||||
     252 AGAACCATTCTGACAAAGGTCCCAAACCAGTTGTCTTCCTGCAACATGGC 301

     100 LeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLe 116
         |||::::::   :::||||||||||:::::||||||   |||:::||||
     302 TTGCTGGCAGATTCTAGTAACTGGGTCACAAACCTTGCCAACAGCAGCCT 351

     116 uGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSerA 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     352 GGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAACAGCA 401

     133 rgGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAsp 149
         ||||||||:::||||||||||||||||||||||||||||:::::||||||
     402 GAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGAT 451

     150 GluPheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeuProAl 166
         ||||||||||||||||||||||||||||||||:::::||||||||||||
     452 GAATTCTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGC 501

     166 aValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrV 183
         |   ||||||||||||||:::||||||||||||||||||:::::|||||||
     502 TTCCATTAACTTCATTCTGAATAAAACTGGCCAAGAACAAGTGTATTATG 551

     183 alGlyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMet 199
         |||||:::|||||||||||||||:::||||||||||||||||||   :::
     552 TGGGTCATTCTCAAGGCACCACTATAGGTTTTATAGCATTTTCACAGATC 601

     200 ProGluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProIleAl 216
         ||||||||||||:::::|||||||||:::|||||||||:::|||:::||
     602 CCTGAGCTGGCTAAAAGGATTAAAATGTTTTTGCCCTGGGTCCTGTGGC 651

     216 aThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeuProA 233
         |:::|||      :::   ||||||   :::|||:::      |||||||
     652 TTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATTACCAG 701

     233 spMetMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThr 249
         ||   :::||||||   |||||||||   ||||||||||||   |||:::
     702 ATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGT 751

     250 ArgPheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAs 266
            |||||||::   |||      ::::::|||   :::|||||||||
     752 GCGTTTTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAA 801

     266 pGlnIleCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnA 283
         :::::::|||:::|||:::   :::||||||   |||||||||   :::|
     802 GGAGCTCTGTGGAAATCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAA 851

     283 snMetAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGly 299
         ||:::|||||||||||||:::::|||||:::::|||:::   ||||||
     852 ATTTAAATATGTCTAGAGTGGATGTATATACAACACATTCTCCTGCTGGA 901

     300 ThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGl 316
         ||||||||||||||:::||||||||||||||||||||:::      ::
     902 ACTTCTGTGCAAAACATGTTACACTGGAGCCAGGCTGTTAAATTCCAAAA 951

     316 uLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysA 333
         :::::::|||||||||||||||||:::::|||||           |
     952 GTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCATTACA 1001

     333 snGlnProThrProValArgTyrArgValArgAspMetThrValProThr 349
         |||||      |||      |||:::|||:::||||||   |||||||||
    1002 ACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACT 1051

     350 AlaMetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLy 366
         |||:::|||:::||||||:::||||||||||:::::      ||||||::
    1052 GCAGTCTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAA 1101

     366 sMetLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIleProG 383
         ::::|||||:::::::::|||||||:::::|||:::::|||||||
    1102 TATCTTACTGACTCAGATCACCAACTTGGTGTTCCATGAGAGCATTCCGG 1151

     383 luTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMet 399
         |||||   |||:::||||||||||||||||||||||||||   |||:::
    1152 AATGGGAGCATCTTGACTTCATTTGGGGCCTGGATGCCCCTTGGAGGCTT 1201

     400 TyrAsnGluIleIleHisLeuMetGlnGln 409
         |||||||:::|||||||:::|||||||:::::
    1202 TATAATAAAATTATTAATCTAATGAGGAAA 1231

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ42310

seq_documentation_block:
```

```
ID   AAQ42310 standard; DNA; 1378 BP.
XX
AC   AAQ42310;
XX
DT   20-SEP-1993  (first entry)
XX
DE   RGL precursor.
XX
KW   Rabbit gastric lipase; RGL; pRGLN2.1; fat; bioconversion;
KW   hydrolysis; transesterification; ds.
XX
OS   Oryctolagus cuniculus.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        120..1247
FT                   /*tag= a
FT                   /note= "claim 1; page 10-11"
XX
PN   EP542629-A.
XX
PD   19-MAY-1993.
XX
PF   12-NOV-1992;   92EP-0403055.
XX
PR   13-NOV-1991;   91FR-0013948.
XX
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Benicourt C,  Blanchard C,  Junien J;
XX
DR   WPI; 1993-161080/20.
DR   P-PSDB; AAR37302.
XX
PT   Rabbit gastric lipase, its precursor and their DNA - useful for
PT   treating conditions linked to gastric lipase deficiency, such as
PT   mucoviscidiosis and pancreatic exocrine insufficiency
XX
PS   Claim 1; Fig 6; 31pp; French.
XX
CC   mRNA was isolated from rabbit gastric mucosal cells, converted to
CC   cDNA and fragments cloned in pUC18.  Recombinant plasmids were used
CC   to transform E.coli MM294 and cells screened for reactivity with
CC   probes corresp. to known parts of the RGL sequence.  One positive
CC   clone contd. plasmid pJ0101 which contd. a 1.35 kb PstI sequence.
CC   This fragment was isolated, subjected to PCR and the amplification
CC   prod. ligated with the 2.2 kb BglII-SacI fragment of pRU276 (contg.
CC   a synthetic Tac promoter and transcription stop signal).  The
CC   recombinant plasmid was designated pRGLN2.1.
XX
SQ   Sequence 1378 BP; 397 A; 287 C; 276 G; 418 T; 0 other;


alignment_scores:
            Quality: 1167.00              Length:    402
              Ratio:  3.505                Gaps:      3
 Percent Similarity:  82.836   Percent Identity: 54.478

alignment_block:
 US-09-578-063-47 x AAQ42310   ..

 Align seg 1/1  to: AAQ42310  from: 1  to: 1378

      16 GluMetTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAs 32
         :::||||||:::|||:::::::||||||   :::::::::      :::::
      51 AAAATGTGGGTGCTTTTCATGGTGGCAGCTTTGCTATCTGCA...CTTGG 97

      32 nSerValHis.................MetProThrLysAlaValAspP 43
           :::::|||                  ||||||          :::|
           :TACACATGGTCTTTTTGGAAAATCAGCACCAACA.........AACC 138

           :luAlaPheMetAsnIleSerGluIleIleGlnHisGlnGlyTyrPro 59
            ||:::   ||||||||||||:::::::|||::::::   |||||||||
           :AAGTGAATATGAATATTAGTCAGATGATTTCCTACTGGGGATACCCA 188

      60 CysGluGluTyrGluValAlaThrGluAspGlyTyrIleLeuSerValAs 76
         |||:::|||||||||:::|||||||||||||||||||||||:::|||||
     189 AGTGAAAAATATGAAGTTGTGACTGAAGATGGTTATATCCTTGAGGTTAA 238

      76 nArgIleProArgGlyLeuValGlnProLysLysThrGlySerArgProV 93
         |||||||||   |||      :::      :::   |||:::|||||||
     239 CAGAATTCCTTATGGGAAGAAAAATTCAGGGAACAGAGGCCAGAGACCCG 288

      93 alValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsnTrpIleSer 109
         |||||:::|||||||||||||||:::::::|||||||||||||||||||
     289 TCGTATTTTTGCAGCATGGTTTGCTTGCATCAGCCTCAAACTGGATCTCC 338

     110 AsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAlaGlyPheAs 126
         ||||||||||||||||||||:::||||||||||||||||||||||:::
     339 AACCTGCCCAACAACAGCCTGGCCTTTATTCTGGCAGATGCTGGTTATGG 388

     126 pValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgLysHisLysT 143
         ||||||:::|||||||||||||||||||:::|||||||||:::::
     389 TGTGTGGCTGGGGAACAGCAGAGGAAATACCTGGTCCAGAAGAAACTTAT 438

     143 hrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMet 159
         |||   |||:::   ||||||||||||||||||:::|||||||||
     439 ACTATTCACCAGACTCCGTTGAATTCTGGGCTTTTAGCTTTGATGAAATG 488

     160 AlaArgPheAspLeuProAlaValIleAsnPheIleLeuGlnLysThrGl 176
         |||:::::|||||||||||:::|||:::||||||::::::::|||||
     489 GCTAAATATGACCTTCCAGCCACAATTGACTTCATTGTAAAGGAAACTGG 538

     176 yGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMetGlyP 193
         |||||||||:::::|||||||||:::|||||||||||||||:::||||
     539 ACAGGAAAAGCTGCACTATGTTGGTCATTCCCAGGGCACCACCATTGGTT 588

     193 heIleAlaPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyr 209
         ||||||||||||||||   |||:::||||||::::::||||||   :::
     589 TTATTGCCTTTTCTACAAATCCCAAGCTGGCTGAAAGAATCAAAACCTTT 638

     210 PheAlaLeuAlaProIleAlaThrValLysHisAlaLysSerProGlyTh 226
         :::|||||||||||||:::||||||||||||:::::||||||      ::
     639 TATGCATTAGCTCCAGTCGCCACTGTGAAGTATACAAAAAGCCTTGTAAA 688

     226 rLysPheLeuLeuLeuProAspMetMetIleLysGlyLeuPheGlyLysL 243
         :|||:::   ::::::|||      |||:::|||   :::||||||   |
     689 CAAACTTAGGTTTATTCCTCCAACCATGTTCAAGATTATATTTGGTGACA 738

     243 ysGluPheLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTyrLeu 259
         ||   |||      ::::::   |||   :::   |||:::      :::
     739 AAATATTCTACCCACACAATTTCTTTGATCAATTTCTTGCCACTCAAGTG 788

     260 CysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMetLeuLeuLe 276
         |||::::::      |||:::   ||||||||||||   ::::::::::
     789 TGTTCCCGTGAGACACTGAATGTAATTTGCAGCAATGCCTTATTTATCAT 838

     276 uGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaSerValTyrA 293
         :   ||||||::::::   |||:::||||||||||||   :::||||||:
     839 TTGTGGCTTTGACAGCGCAAACTTGAACATGAGTCGCTTGGATGTGTACG 888

     293 laAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrpSer 309
         :::::|||:::   ||||||||||||||||||||:::||||||||||:::
     889 TGTCACATAATCCGGCAGGAACTTCAGTTCAAAACATGCTGCACTGGACC 938

     310 GlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGlySerGluTh 326
         |||||||||:::||||||:::::::::||||||:::|||||||   ::
     939 CAGGCTGTTAAATCTGGAAATTTTCAAGCTTTTAATTGGGGAAGTCCAGC 988

     326 rLysAsnLeuGluLysCysAsnGlnProThrProValArgTyrArgValA 343
         ::::|||:::         ||||||||||||||      |||:::|||
     989 TCAGAATGTAGTGCACTTCAATCAGCCCACACCTCCCTACTACAATGTGA 1038

     343 rgAspMetThrValProThrAlaMetTrpThrGlyGlyGlnAspTrpLeu 359
             |||:::||||||   |||:::|||:::||||||:::|||||||||
    1039 CCGCCATGAATGTGCCAATTGCAGTGTGGAGTGGTGGCAATGACTGGTTG 1088
```

```
     360 SerAsnProGluAspValLysMetLeuLeuSerGluValThrAsnLeuIl 376
         ::::::|||:::|||||   :::||||||   :::::::::|||||||
    1089 GCTGACCCCCAAGATGTTGACCTTTTGCTTCCAAAACTTTCTAATCTCAT 1138

     376 eTyrHisLysAsnIleProGluTrpAlaHisValAspPheIleTrpGlyL 393
         ||||||||||:::|||      :::   |||:::||||||||||||:::: 
    1139 TTACCACAAGGAGATTCTTCCATATAATCACTTGGATTTTATCTGGGCAA 1188

     393 euAspAlaProHisArgMetTyrAsnGluIleIleHisLeuMetGlnGln 409
         :::::||||||:::::::::|||||||||||||||   :::|||   ::: 
    1189 TGAATGCTCCTCAAGAAGTTTACAATGAAATTATTTCTATGATGGCAAAA 1238

     410 GluGlu 411
         ::::::
    1239 GATAAA 1244

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:AAN50385

seq_documentation_block:
ID   AAN50385 standard; DNA; 1336 BP.
XX
AC   AAN50385;
XX
DT   17-JAN-1992  (first entry)
XX
DE   Rat prelingual lipase gene.
XX
KW   Prelingual lipase; enzyme; EC-3.1.1.3; ss.
XX
OS   Rattus rattus.
XX
FH   Key             Location/Qualifiers
FT   CDS             7..1194
FT                   /*tag= a
XX
PN   GB2142337-A.
XX
PD   16-JAN-1985.
XX
PF   29-JUN-1984;   84GB-0016581.
XX
PR   01-JUL-1983;   83GB-0017989.
PR   05-SEP-1983;   83GB-0023759.
XX
PA   (CELL-) CELLTECH LTD.
XX
PI   Carey NH,  Williamson R;
XX
DR   WPI; 1985-014450/03.
DR   P-PSDB; AAP50322.
XX
PT   New lingual lipase protein for treatment of lipase deficiency - also
PT   new pre-lingual lipase protein and related products
XX
PS   Disclosure; Fig 5; 15pp; English.
XX
CC   This sequence may be expressed in a transformant host organism using
CC   a vector plasmid in order to produce a prelingual lipase protein.
CC   The expressed protein may be used for the treatment of lipase
CC   deficiency.
XX
SQ   Sequence 1336 BP; 409 A; 285 C; 263 G; 379 T; 0 other;


alignment_scores:
            Quality: 1158.50            Length:    399
              Ratio:  3.543              Gaps:      1
 Percent Similarity:  81.955   Percent Identity: 54.386

alignment_block:
 US-09-578-063-47 x AAN50385   ..

 Align seg 1/1  to: AAN50385  from: 1  to: 1336
```

```
  16 GluMetTrpLeuLeuIleLeuValAlaTyrMetPheGln...ArgAsnVa 31
     :::|||||||||||:::::::::::   :::             :::::
   4 AAGATGTGGCTGTTATTAATAACAAGTGTGATATCAACATTCGGAGGTGC 53

  31 lAsnSerValHisMetProThrLysAlaValAspProGluAlaPheMetA 48
     ::::::::::                  :::|||||||||   ||||
  54 ACATGGCCTATTTGGAAAACTGGGTCCTGGAAACCCTGAAGCAAATATGA 103

  48 snIleSerGluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGlu 64
     ||||||||:::::|||   :::   ||||||||||||:::|||||||||
 104 ATATTAGTCAGATGATAACTTACTGGGGATATCCATGTCAAGAATATGAA 153

  65 ValAlaThrGluAspGlyTyrIleLeuSerValAsnArgIleProArgGl 81
     |||:::|||||||||||||||||||||||:::|||   ||||||||||:::||
 154 GTTGTTACTGAAGATGGCTACATTCTGGGGGTCTACAGAATTCCTCATGG 203

  81 yLeuValGlnProLysLysThrGlySerArgProValValLeuLeuGlnH 98
     |      :::   ::::::   |||:::||||||||||||   |||||||
 204 GAAGAATAATTCTGAAAATATAGGCAAGAGACCTGTGGTGTATTTGCAGC 253

  98 isGlyLeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsn 114
     ||||||||:::::::::|||:::|||||||||:::|||||||||||||||
 254 ATGGTTTGATCGCATCAGCCACAAACTGGATTGCAAATCTACCAAACAAC 303

 115 SerLeuGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAs 131
     ||||||:::|||:::|||||||||||||||:::|||||||||:::||||||
 304 AGCCTCGCCTTTATGCTAGCAGACGCTGGCTATGATGTGTGGCTGGGGAA 353

 131 nSerArgGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspG 148
     ||||||||||||:::||||||||||||:::          |||   |||:
 354 CAGTCGAGGAAATACATGGTCCAGGAAAAATGTATACTACTCACCAGACT 403

 148 lnAspGluPheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeu 164
     ::   ||||||||||||||||||:::||||||||||||:::::::||||||
 404 CAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAATATGACCTT 453

 165 ProAlaValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTy 181
     ||||||:::|||||||||||:::|||||||||||||||||||||||||::
 454 CCCGCCACAATAAACTTCATTGTACAGAAAACTGGACAAGAGAAGATACA 503

 181 rTyrValGlyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerT 198
     :|||||||||:::||||||   ||||||:::|||||||||||||||||||
 504 CTATGTTGGTCATTCTCAGCGCACCACTATTGGTTTCATTGCCTTTTCTA 553

 198 hrMetProGluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaPro 214
     ||   |||   ||||||:::||||||||   ::::::||||||||||||
 554 CCAATCCTACACTGGCCAAAAAAATCAAGACGTTTTATGCATTAGCTCCA 603

 215 IleAlaThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuLe 231
     :::||||||||||||:::::::::||||||      |||:::   :::::
 604 GTTGCTACCGTGAAGTATACACAAAGTCCCTTGAAAAAGATTTCATTTAT 653

 231 uProAspMetMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrG 248
     :|||   ::::::::::|||   :::||||||||||||   ||||||   :
 654 TCCTACATTTCTTTTCAAGCTTATGTTTGGCAAGAAAATGTTCCTGCCCC 703

 248 lnThrArgPheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIle 264
     ::|||   |||         |||         :::|||::::::   :::
 704 ACACCTACTTTGATGACTTTCTTGGTACCGAAGTGTGCTCACGGGAGGTT 753

 265 LeuAspGlnIleCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnTh 281
     ||||||   :::||||||||||   ::::::::::::   |||||||:::
 754 CTAGATCTTCTCTGCAGCAACACTTTATTCATCTTCTGTGGATTTGACAA 803

 281 rAsnAsnMetAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuA 298
      :::|||:::|||:::||||||   :::||||||   :::|||:::   |
 804 GAAAAACTTAAATGTGAGTCGTTTTGATGTGTATCTAGGGCATAATCCAG 853

 298 laGlyThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSer 314
     ||||||||||||||||||::::::||||||||:::|||   |||:::|||
 854 CAGGGACATCTGTTCAAGACTTTCTCCACTGGGCACAGCTTGTTAGATCT 903
```

```
     315 GlyGluLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLy 331
     904 GGGAAATTTCAAGCCTTCAACTGGGGAAGCCCATCCCAGAACATGTTACA 953

     331 sCysAsnGlnProThrProValArgTyrArgValArgAspMetThrValP 348
     954 CTACAACCAGAAAACGCCTCCTGAATATGATGTGTCAGCCATGACTGTGC 1003

     348 roThrAlaMetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAsp 364
    1004 CAGTTGCAGTGTGGAACGGTGGCAATGACATCCTGGCTGATCCCCAAGAT 1053

     365 ValLysMetLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIl 381
    1054 GTCGCCATGCTGCTTCCCAAACTCTCCAACCTCCTGTTCCATAAGGAGAT 1103

     381 eProGluTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisA 398
    1104 TCTTGCCTACAATCACCTGGACTTCATCTGGGCAATGGATGCCCCTCAAG 1153

     398 rgMetTyrAsnGluIleIleHisLeuMetGlnGlnGluGluThrAsn 413
    1154 AGGTTTACAATGAGATGATTTCCATGATGGCAGAAGACTAAAAGAAT 1200

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ68388

seq_documentation_block:
ID   AAQ68388 standard; DNA; 1528 BP.
XX
AC   AAQ68388;
XX
DT   20-FEB-1995  (first entry)
XX
DE   Canine gastric lipase.
XX
KW   Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
KW   bioconversion; exocrine pancreatic insufficiency; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1140
FT                   /*tag= a
FT   misc_RNA        1..1137
FT                   /*tag= b
FT                   /note= "claim 1"
XX
PN   WO9413816-A.
XX
PD   23-JUN-1994.
XX
PF   16-DEC-1993;   93WO-FR01260.
XX
PR   16-DEC-1992;   92FR-0015201.
XX
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Benicourt C,  Blanchard C,  Junien J;
XX
DR   WPI; 1994-217890/26.
DR   P-PSDB; AAR56870.
XX
PT   Recombinant canine gastric lipase and nucleic acid encoding it -
PT   are used for improving absorption of ingested fat, treating
PT   mucoviscidosis etc. and in enzymatic bio-conversions
XX
PS   Claim 1; Fig 8; 52pp; French.
XX
CC   The sequence given below is the sequence of figure 8, altered
CC   according to the amendments described on page 2 of the appended
CC   letter.
CC   CGL is used to improve absorption of ingested fat, in healthy and
CC   sick patients (e.g. having altered levels of gastric lipase); to
CC   treat conditions associated with insufficiency (or lack) of lipases,
CC   esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC   where immobilised, for bioconversions, e.g. hydrolysis or
CC   transesterification (other mammalian gastric lipases, or derivs.,
CC   can be used in this application).
XX
SQ   Sequence 1528 BP; 394 A; 386 C; 328 G; 420 T; 0 other;


alignment_scores:
           Quality: 1141.50            Length:    372
             Ratio:  3.624              Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
 US-09-578-063-47 x AAQ68388   ..

 Align seq 1/1  to: AAQ68388  from: 1  to: 1528

      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
      19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
      60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

      70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
     110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

      87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
     160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

     103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
     210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

     120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
     260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

     137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
     310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 359

     153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
     360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409

     170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
     410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

     187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
     460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

     203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
     510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

     220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
     560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

     237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
     610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

     253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
```

```
     657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

     269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
         ||||||||   ::::::::::::   |||||||::|||   |||:::|||
     707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

     286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
         |||||||||   :::||||||   :::|||:::   ||||||||||||||
     757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

     302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
         |||||||:::||||||||||||||||||||||:::||||||:::::::::|
     807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

     319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
         ||||||||||||||||   :::::::|||:::          :::|||
     857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

     336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
         |||      |||::::::   ||||||   ||||||   |||:::||
     907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

     352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
         |:::|||||:::|||   |||:::::::|||:::|||||   :::||||
     957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

     369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
         |||||:::::::   ||||||||||||||||:::::|||||   :::
    1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

     386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
         |||:::||||||||||||:::::||||||||||:::   :::||||||||
    1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

     402 uIleIleHisLeuMet 407
         ||||:::   :::|||
    1107 AATTGTTTCCATGATG 1122
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT58915

seq_documentation_block:

```
ID   AAT58915 standard; cDNA to mRNA; 1528 BP.
XX
AC   AAT58915;
XX
DT   19-AUG-1997  (first entry)
XX
DE   Dog gastric lipase coding sequence.
XX
KW   Dog; duodenal; gastric; lipase; transgenic; plant; recombinant; extract;
KW   food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;
KW   hydrolysis; trans-esterification; substrate; enzyme; biofuel; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1140
FT                   /*tag= a
FT                   /product= gastric lipase
XX
PN   WO9633277-A2.
XX
PD   24-OCT-1996.
XX
PF   19-APR-1996;   96WO-FR00606.
XX
PR   20-APR-1995;   95FR-0004754.
XX
PA   (BIOC-) BIOCEM SA.
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Baudino S,  Benicourt C,  Cudrey C,  Gruber V,  Lenee P;
PI   Merot B;
XX
DR   WPI; 1996-485783/48.
DR   P-PSDB; AAW09382.
XX
PT   Recombinant human or canine pre-duodenal lipase prodn. in transgenic
PT   plants - useful for facilitating absorption of fat, as bio-catalysts
PT   and for prodn. of bio-fuel
XX
PS   Claim 2; Fig 1; 130pp; French.
XX
CC   This is the nucleotide sequence encoding the dog pre-duodenal (i.e.
CC   gastric) lipase enzyme.  The sequence can be used to generate transgenic
CC   plants producing recombinant lipase in an enzymatically active form.
CC   Alternatively the first 12 bases of this sequence (encoding amino acids
CC   1-4 of the protein encoded by this sequence) or bases 1-162 of this
CC   sequence (encoding amino acids 1-54) can be deleted to form the
CC   derivatives designated delta-4 or delta-54 respectively.  Plants, or
CC   their extracts, expressing the lipases or the truncated derivatives, can
CC   be used: (a) as pharmaceuticals or food to facilitate absorption of fat,
CC   either in healthy subjects or in patients with inadequate levels of
CC   gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine
CC   pancreatic insufficient, in the elderly or in patients undergoing medical
CC   treatment which alters fat adsorption; (b) for performing industrial or
CC   agricultural reactions, e.g. in processing of fats or in the dairy
CC   industry, for hydrolysis or transesterification reactions, etc., where
CC   the plant material may provide both enzyme and substrate.  The transgenic
CC   plants can also be used for biofuel production.
XX
SQ   Sequence 1528 BP; 392 A; 387 C; 329 G; 420 T; 0 other;
```

alignment_scores:

```
            Quality: 1141.50              Length:    372
              Ratio:  3.624                 Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452
```

alignment_block:

US-09-578-063-47 x AAT58915 ..

Align seg 1/1 to: AAT58915 from: 1 to: 1528

```
      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
         ||||||          :::|||||:::   |||||||||||:::::||
      19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
         |   :::   ||||||||||:::|||||||||||||||:::||||||||||
      60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

      70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
         ||||||||||:::::::::|||||||||   |||       :::   :::
     110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

      87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
         :::   |||   ||||||||:::::|||||||||||||||||:::::::::
     160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

     103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
         :|||:::|||||||||||||||||||||||||||||||||:::|||||||
     210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

     120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
         |||||||||||||:::|||||||||:::|||||||||||||||||:::
     260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

     137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
         |||:::|||::::::          |||   |||:::   ||||||||||
     310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 359

     153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
         |||||||:::|||||||||||||:::::::|||||||||||||:::|||:::|
     360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409
```

```
    170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
    410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

    187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
    460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

    203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
    510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

    220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
    560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

    237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
    610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

    253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
    657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

    269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
    707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

    286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
    757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

    302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
    807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

    319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
    857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

    336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
    907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

    352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
    957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

    369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
   1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

    386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
   1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

    402 uIleIleHisLeuMet 407
   1107 AATTGTTTCCATGATG 1122

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ68389

seq_documentation_block:
ID   AAQ68389 standard; DNA; 1531 BP.
XX
AC   AAQ68389;
XX
DT   20-FEB-1995  (first entry)
XX
DE   Canine gastric lipase (Met-CGL).
XX
KW   Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
KW   bioconversion; exocrine pancreatic insufficiency; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1143
FT                   /*tag= a
XX
PN   WO9413816-A.
XX
PD   23-JUN-1994.
XX
PF   16-DEC-1993;   93WO-FR01260.
XX
PR   16-DEC-1992;   92FR-0015201.
XX
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Benicourt C,  Blanchard C,  Junien J;
XX
DR   WPI; 1994-217890/26.
DR   P-PSDB; AAR56871.
XX
PT   Recombinant canine gastric lipase and nucleic acid encoding it -
PT   are used for improving absorption of ingested fat, treating
PT   mucoviscidosis etc. and in enzymatic bio-conversions
XX
PS   Claim 2; Fig 8; 52pp; French.
XX
CC   The sequence given below is the sequence of figure 8, altered
CC   according to the amendments described on page 2 of the appended
CC   letter.
CC   CGL is used to improve absorption of ingested fat, in healthy and
CC   sick patients (e.g. having altered levels of gastric lipase); to
CC   treat conditions associated with insufficiency (or lack) of lipases,
CC   esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC   where immobilised, for bioconversions, e.g. hydrolysis or
CC   transesterification (other mammalian gastric lipases, or derivs.,
CC   can be used in this application).
XX
SQ   Sequence 1531 BP; 395 A; 386 C; 329 G; 421 T; 0 other;


alignment_scores:
            Quality: 1141.50              Length:    372
              Ratio:  3.624                Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
 US-09-578-063-47 x AAQ68389   ..

 Align seg 1/1  to: AAQ68389  from: 1  to: 1531

      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
      22 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 62

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
      63 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 112

      70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
     113 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 162

      87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
     163 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 212

     103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
     213 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 262

     120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
```

```
263 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 312

137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
313 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 362

153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
363 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 412

170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
413 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 462

187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
463 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 512

203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
513 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 562

220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
563 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 612

237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
613 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 659

253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
660 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 709

269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
710 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 759

286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
760 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 809

302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
810 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 859

319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
860 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 909

336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
910 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 959

352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
960 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1009

369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
1010 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1059

386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
1060 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1109

402 uIleIleHisLeuMet 407
1110 AATTGTTTCCATGATG 1125
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:AAN60685

```
seq_documentation_block:
ID   AAN60685 standard; DNA; 1367 BP.
XX
AC   AAN60685;
XX
DT   08-JUN-1991  (first entry)
XX
DE   Sequence encoding pregastric lipase.
XX
KW   Precursor polypeptide; secretion vector; enzyme; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     47..103
FT                   /*tag= a
FT   mat_peptide     104..1243
FT                   /*tag= b
XX
PN   WO8603778-A.
XX
PD   03-JUL-1986.
XX
PF   23-DEC-1985;   85WO-GB00599.
XX
PR   21-DEC-1984;   84GB-0032483.
PR   23-DEC-1985;   85WO-GB00599.
PR   01-JAN-1986;   86GB-0019568.
XX
PA   (BREW-) BREWING RES FOUND.
PA   (TUBB/) TUBB R S.
XX
PI   Tubb RS;
XX
DR   WPI; 1986-182910/28.
DR   P-PSDB; AAP60724.
XX
PT   New precursor polypeptide of defined sequence - and corresp. DNA
PT   used to transform hosts for prodn. of the polypeptide
XX
PS   Example; Fig 5; 60pp; English.
XX
CC   The inventors claim a DNA sequence encoding AAP60723 linked to a
CC   promoter upstream and a gene for a polypeptide downstream.
CC   Particular examples are the yeast emzyme AMG, the mammalian enzyme,
CC   gastric lipase and the mammalian lymphokine, interferon-alpha2.
XX
SQ   Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;


alignment_scores:
            Quality: 1124.00              Length:    405
              Ratio:  3.416                Gaps:      2
 Percent Similarity:  81.235   Percent Identity: 52.099

alignment_block:
 US-09-578-063-47 x AAN60685   ..

 Align seg 1/1  to: AAN60685  from: 1  to: 1367

      14 ArgMetGluMetTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAs 30
      38 AGGTCCAAAATGTGGCTGCTTTTAACAATGGCAAGTTTGATA..... 79

      30 nValAsnSerValHisMetProThrLysAlaVal..... 41
      80 .....TCTGTACTGGGGACTACACATGGTTTGTTTGGAAAATTACATC 122

      42 .....AspProGluAlaPheMetAsnIleSerGluIleIleGlnHisGln 56
     123 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGG 172

      57 GlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIleLe 73
```

```
     |||||||||  ||||||||||||||:::|||||||||||||||||||
 173 GGATACCCAAATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCT 222

  73 uSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrGlyS 90
     |:::|||||||||||||||  |||      :::      :::||||||:
 223 TGAAGTCAATAGAATTCCTTATGGGAAGAAAAATTCAGGGAATACAGGCC 272

  90 erArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsn 106
     ::|||||||||||:::|||||||||||||||:::::::::|||:::|||
 273 AGAGACCTGTTGTGTTTTTGCAGCATGGTTTGCTTGCATCAGCCACAAAC 322

 107 TrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAl 123
     ||||||||||||||||||||||||||||:::||||||||||||||||||
 323 TGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTCTGGCAGATGC 372

 123 aGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgL 140
     ||||:::||||||||||:::|||||||||||||||||||:::|||:::|||:
 373 TGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 422

 140 ysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyr 156
     :::::         |||   |||:::   ||||||||||||||||||:::
 423 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTT 472

 157 AspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLeuGl 173
     ||||||||||||:::::||||||||||||:::|||:::||||||:::::
 473 GATGAAATGGCTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAA 522

 173 nLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrT 190
     :|||||||||||:::::::::::::||||||||:::|||||||||||||
 523 GAAAACTGGACAGAAGCAGCTACACTATGTTGGCCATTCCCAGGGCACCA 572

 190 hrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLysIle 206
     ||:::|||||||||||||||||||||   |||:::||||||:::::|||
 573 CCATTGGTTTTATTGCCTTTTCCACCAATCCCAGCCTGGCTAAAAGAATC 622

 207 LysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLysSe 223
     |||   ::::::||||||||||||:::||||||||||||::::::|||||
 623 AAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGTATACAAAAAG 672

 223 rProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyLeuP 240
     |     :::|||:::   :::::::|||:::   :::::::|||   :::|
 673 CCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 722

 240 heGlyLysLysGluPheLeuTyrGlnThrArgPheLeuArgGlnLeuVal 256
     |||||   |||   |||       ::::::   |||   :::   |||:::
 723 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCT 772

 257 IleTyrLeuCysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMe 273
           :::|||::::::   :::|||:::   :::|||||||||   ::
 773 ACTGAAGTGTGCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTT 822

 273 tLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaS 290
     ::::::::::   ||||||:::::::::|||:::|||   ||||||   :
 823 ATTTATAATTTGTGGATTTGACAGTAAGAACTTTAACACGAGTCGCTTGG 872

 290 erValTyrAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeu 306
     ::||||||   :::|||:::   ||||||||||||||||||||:::::::
 873 ATGTGTATCTATCACATAATCCAGCAGGAACTTCTGTTCAAAACATGTTC 922

 307 HisTrpSerGlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGl 323
     ||||||:::||||||||:::||||||:::::::::|||:::||||||||
 923 CATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTTATGACTGGGG 972

 323 ySerGluThrLysAsnLeuGluLysCysAsnGlnProThrProValArgT 340
     ||||  ::::::|||           :::|||     |||      |
 973 AAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1022

 340 yrArgValArgAspMetThrValProThrAlaMetTrpThrGlyGlyGln 356
     ||:::|||     |||:::||||||   |||:::|||:::||||||:::
1023 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAG 1072

 357 AspTrpLeuSerAsnProGluAspValLysMetLeuLeuSerGluValTh 373
     |||   |||::::::|||:::||||||   :::|||||||   ::::::
```

```
1073 GACCTGTTGGCTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCC 1122

 373 rAsnLeuIleTyrHisLysAsnIleProGluTrpAlaHisValAspPheI 390
      ||||||||||||||||||:::||||||   :::   |||:::|||||||
1123 CAATCTTATTTACCACAAGGAGATTCCTTTTTACAATCACTTGGACTTTA 1172

 390 leTrpGlyLeuAspAlaProHisArgMetTyrAsnGluIleIleHisLeu 406
      |||||::::::|||||||||:::::::::|||||:::|||:::   :::
1173 TCTGGGCAATGGATGCCCCTCAAGAAGTTTACAATGACATTGTTTCTATG 1222

 407 MetGlnGlnGluGlu 411
      :::::::::::::::
1223 ATATCAGAAGATAAA 1237
```

```
seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:AAN60566

seq_documentation_block:
ID   AAN60566 standard; DNA; 1367 BP.
XX
AC   AAN60566;
XX
DT   22-AUG-1991  (first entry)
XX
DE   Sequence encoding human pregastric lipase.
XX
KW   Cystic fibrosis therapy; enzyme; lipase deficiency; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             47..103
FT                   /*tag= a
FT   mat_peptide     104..1243
FT                   /*tag= b
XX
PN   WO8601532-A.
XX
PD   13-MAR-1986.
XX
PF   15-AUG-1985;   85WO-GB00364.
XX
PR   21-AUG-1984;   84GB-0021210.
PR   15-AUG-1985;   85WO-GB00364.
PR   01-JAN-1986;   86GB-0008897.
XX
PA   (CELL-) CELLTECH LTD.
PA   (LOWE/) LOWE P A.
XX
PI   Lowe PA;
XX
DR   WPI; 1986-081634/12.
DR   P-PSDB; AAP60658.
XX
PT   New gastric lipase protein, esp. of human origin - for treating
PT   lipase deficiency, and DNA sequences coding for it
XX
PS   Disclosure; Fig 3; 39pp; English.
XX
CC   The inventors claim a pregastric lipase protein and a gene encoding
CC   it. Gastric lipase protein is useful for oral administration to
CC   treat lipase deficiency, e.g. cystic fibrosis or pancreatitis.
XX
SQ   Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;


alignment_scores:
           Quality: 1124.00            Length:    405
             Ratio:  3.416              Gaps:      2
 Percent Similarity:  81.235   Percent Identity: 52.099

alignment_block:
 US-09-578-063-47 x AAN60566   ..
```

```
Align seg 1/1  to: AAN60566  from: 1  to: 1367

     14 ArgMetGluMetTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAs 30
        |||   :::|||||||||||||:::   :::|||   ::::::
     38 AGGTCCAAAATGTGGCTGCTTTTAACAATGGCAAGTTTGATA........ 79

     30 nValAsnSerValHisMetProThrLysAlaVal............... 41
             ||||||          |||   ::::::
     80 .......TCTGTACTGGGGACTACACATGGTTTGTTTGGAAAATTACATC 122

     42 .....AspProGluAlaPheMetAsnIleSerGluIleIleGlnHisGln 56
           :::||||||:::   ||||||||||||:::::|||   :::
    123 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGG 172

     57 GlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIleLe 73
        ||||||||   ||||||||||||||:::||||||||||||||||||||
    173 GGATACCCAAATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCT 222

     73 uSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrGlyS 90
        |:::|||||||||||||||   |||      :::      :::||||||:
    223 TGAAGTCAATAGAATTCCTTATGGGAAGAAAAATTCAGGGAATACAGGCC 272

     90 erArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsn 106
        ::|||||||||||:::|||||||||||||||||:::::::|||:::|||
    273 AGAGACCTGTTGTGTTTTTGCAGCATGGTTTGCTTGCATCAGCCACAAAC 322

    107 TrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAl 123
        ||||||||||||||||||||||||||||||:::|||||||||||||||||
    323 TGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTCTGGCAGATGC 372

    123 aGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgL 140
        ||||:::|||||||||:::|||||||||||||||||||:::|||:::|||:
    373 TGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 422

    140 ysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyr 156
        :::::          |||   |||:::   |||||||||||||||||:::
    423 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTT 472

    157 AspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLeuGl 173
        ||||||||||||:::::|||||||||||||:::|||:::||||||:::::
    473 GATGAAATGGCTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAA 522

    173 nLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrT 190
        :|||||||||||||:::::::::::::|||||||||:::|||||||||||||
    523 GAAAACTGGACAGAAGCAGCTACACTATGTTGGCCATTCCCAGGGCACCA 572

    190 hrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLysIle 206
        ||:::|||||||||||||||||||||   |||:::||||||:::::::|||
    573 CCATTGGTTTTATTGCCTTTTCCACCAATCCCAGCCTGGCTAAAAGAATC 622

    207 LysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLysSe 223
        |||   ::::::||||||||||||:::|||||||||||||:::::::|||||
    623 AAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGTATACAAAAAG 672

    223 rProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyLeuP 240
        |     :::|||:::   ::::::|||:::   ::::::|||   :::|
    673 CCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 722

    240 heGlyLysLysGluPheLeuTyrGlnThrArgPheLeuArgGlnLeuVal 256
        |||||   |||   |||      ::::::   |||   :::   |||:::
    723 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCT 772

    257 IleTyrLeuCysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMe 273
              :::|||::::::   :::|||:::   :::|||||||||   ::
    773 ACTGAAGTGTGCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTT 822

    273 tLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaS 290
        ::::::::::   ||||||:::::::::|||:::|||   ||||||   :
    823 ATTTATAATTTGTGGATTTGACAGTAAGAACTTTAACACGAGTCGCTTGG 872

    290 erValTyrAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeu 306
        ::||||||   :::|||:::   ||||||||||||||||||||||:::::
    873 ATGTGTATCTATCACATAATCCAGCAGGAACTTCTGTTCAAAACATGTTC 922

    307 HisTrpSerGlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGl 323
        ||||||:::|||||||||:::||||||:::::::::|||:::|||||||||
    923 CATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTTATGACTGGGG 972

    323 ySerGluThrLysAsnLeuGluLysCysAsnGlnProThrProValArgT 340
        ||||   ::::::|||            :::|||      |||      |
    973 AAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1022

    340 yrArgValArgAspMetThrValProThrAlaMetTrpThrGlyGlyGln 356
        ||:::|||      |||:::||||||   |||:::|||:::||||||:::
   1023 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAG 1072

    357 AspTrpLeuSerAsnProGluAspValLysMetLeuLeuSerGluValTh 373
        |||   |||:::::::|||:::||||||   :::||||||   ::::::
   1073 GACCTGTTGGCTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCC 1122

    373 rAsnLeuIleTyrHisLysAsnIleProGluTrpAlaHisValAspPheI 390
         ||||||||||||||||||:::||||||   :::   |||:::|||||||
   1123 CAATCTTATTTACCACAAGGAGATTCCTTTTTACAATCACTTGGACTTTA 1172

    390 leTrpGlyLeuAspAlaProHisArgMetTyrAsnGluIleIleHisLeu 406
        |||||::::::||||||||||:::::::::|||||||:::|||:::   :::
   1173 TCTGGGCAATGGATGCCCCTCAAGAAGTTTACAATGACATTGTTTCTATG 1222

    407 MetGlnGlnGluGlu 411
        :::::::::::::::
   1223 ATATCAGAAGATAAA 1237

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT58916

seq_documentation_block:
ID   AAT58916 standard; DNA; 1367 BP.
XX
AC   AAT58916;
XX
DT   19-AUG-1997  (first entry)
XX
DE   Human gastric lipase coding sequence.
XX
KW   Duodenal; gastric; lipase; transgenic; plant; recombinant; extract; ss;
KW   food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;
KW   hydrolysis; trans-esterification; substrate; enzyme; biofuel; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             47..1243
FT                   /*tag= a
FT                   /product= gastric lipase
FT   sig_peptide     47..103
FT                   /*tag= b
FT   mat_peptide     104..1240
FT                   /*tag= c
XX
PN   WO9633277-A2.
XX
PD   24-OCT-1996.
XX
PF   19-APR-1996;   96WO-FR00606.
XX
PR   20-APR-1995;   95FR-0004754.
XX
PA   (BIOC-) BIOCEM SA.
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Baudino S,  Benicourt C,  Cudrey C,  Gruber V,  Lenee P;
PI   Merot B;
XX
DR   WPI; 1996-485783/48.
DR   P-PSDB; AAW09383.
XX
PT   Recombinant human or canine pre-duodenal lipase prodn. in transgenic
```

```
PT   plants - useful for facilitating absorption of fat, as bio-catalysts
PT   and for prodn. of bio-fuel
XX
PS   Claim 3; Fig 4; 130pp; French.
XX
CC   This is the nucleotide sequence encoding the human pre-duodenal (i.e.
CC   gastric) lipase enzyme.  The sequence can be used to generate transgenic
CC   plants producing recombinant lipase in an enzymatically active form.
CC   Alternatively bases encoding amino acids 20-23 or 20-73 (1-4 or 1-54 of
CC   the mature protein, respectively) can be deleted to form the derivatives
CC   designated delta-4 or delta-54 respectively.  Plants, or their extracts,
CC   expressing the lipases or the truncated derivatives, can be used:
CC   (a) as pharmaceuticals or food to facilitate absorption of fat, either
CC   in healthy subjects or in patients with inadequate levels of
CC   gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine
CC   pancreatic insufficient, in the elderly or in patients undergoing medical
CC   treatment which alters fat adsorption; (b) for performing industrial or
CC   agricultural reactions, e.g. in processing of fats or in the dairy
CC   industry, for hydrolysis or transesterification reactions, etc., where
CC   the plant material may provide both enzyme and substrate.  The transgenic
CC   plants can also be used for biofuel production.
XX
SQ   Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;


alignment_scores:
           Quality: 1124.00              Length:    405
             Ratio:  3.416                Gaps:      2
 Percent Similarity:  81.235   Percent Identity: 52.099

alignment_block:
 US-09-578-063-47 x AAT58916   ..

 Align seg 1/1  to: AAT58916  from: 1  to: 1367

      14 ArgMetGluMetTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAs 30
         |||   :::|||||||||||||:::   :::|||   ::::::
      38 AGGTCCAAAATGTGGCTGCTTTTAACAATGGCAAGTTTGATA..... 79

      30 nValAsnSerValHisMetProThrLysAlaVal.............. 41
                |||||          |||   ::::::
      80 .......TCTGTACTGGGGACTACACATGGTTTGTTTGGAAAATTACATC 122

      42 .....AspProGluAlaPheMetAsnIleSerGluIleIleGlnHisGln 56
           :::|||||:::   ||||||||||||:::::|||   :::
     123 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGG 172

      57 GlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIleLe 73
         |||||||||   ||||||||||||||:::|||||||||||||||||||
     173 GGATACCCAAATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCT 222

      73 uSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrGlyS 90
         |:::|||||||||||||||   |||      :::      :::|||||||:
     223 TGAAGTCAATAGAATTCCTTATGGGAAGAAAAATTCAGGGAATACAGGCC 272

      90 erArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsn 106
         ::|||||||||||:::|||||||||||||||:::::::::|||:::|||
     273 AGAGACCTGTTGTGTTTTTGCAGCATGGTTTGCTTGCATCAGCCACAAAC 322

     107 TrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAl 123
         ||||||||||||||||||||||||||||||:::||||||||||||||||
     323 TGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTCTGGCAGATGC 372

     123 aGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgL 140
         ||||:::|||||||||:::|||||||||||||||||||:::|||:::|||:
     373 TGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 422

     140 ysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyr 156
         :::::          |||   |||:::   ||||||||||||||||::
     423 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTT 472

     157 AspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLeuGl 173
         ||||||||||||:::::|||||||||||||:::|||:::|||||:::::

     473 GATGAAATGGCTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAA 522

     173 nLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrT 190
         :|||||||||||||:::::::::::::|||||||||::::|||||||||||||
     523 GAAAACTGGACAGAAGCAGCTACACTATGTTGGCCATTCCCAGGGCACCA 572

     190 hrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLysIle 206
         ||:::|||||||||||||||||||||   |||:::||||||:::::|||
     573 CCATTGGTTTTATTGCCTTTTCCACCAATCCCAGCCTGGCTAAAAGAATC 622

     207 LysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLysSe 223
         |||   ::::::|||||||||||||:::|||||||||||||:::::|||||
     623 AAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGTATACAAAAAG 672

     223 rProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyLeuP 240
         |      :::|||:::   ::::::|||:::   :::::::||   :::|
     673 CCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 722

     240 heGlyLysLysGluPheLeuTyrGlnThrArgPheLeuArgGlnLeuVal 256
         |||||   |||   |||      ::::::   |||   :::   |||:::
     723 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCT 772

     257 IleTyrLeuCysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMe 273
             :::|||:::::   :::|||:::   :::|||||||||   ::
     773 ACTGAAGTGTGCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTT 822

     273 tLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaS 290
         ::::::::::   ||||||:::::::::|||:::|||   ||||||   :
     823 ATTTATAATTTGTGGATTTGACAGTAAGAACTTTAACACGAGTCGCTTGG 872

     290 erValTyrAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeu 306
         ::||||||   :::|||:::   ||||||||||||||||||||:::::
     873 ATGTGTATCTATCACATAATCCAGCAGGAACTTCTGTTCAAAACATGTTC 922

     307 HisTrpSerGlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGl 323
         ||||||:::||||||||||:::||||||:::::::::|||:::|||||||
     923 CATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTTATGACTGGGG 972

     323 ySerGluThrLysAsnLeuGluLysCysAsnGlnProThrProValArgT 340
         ||||   :::::::|||            ::|||      |||      |
     973 AAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1022

     340 yrArgValArgAspMetThrValProThrAlaMetTrpThrGlyGlyGln 356
         ||:::|||      |||:::||||||   |||:::|||:::||||||:::
    1023 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAG 1072

     357 AspTrpLeuSerAsnProGluAspValLysMetLeuLeuSerGluValTh 373
         |||   |||:::::|||:::||||||   :::||||||   ::::::
    1073 GACCTGTTGGCTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCC 1122

     373 rAsnLeuIleTyrHisLysAsnIleProGluTrpAlaHisValAspPheI 390
          ||||||||||||||||||:::||||||   :::   |||:::|||||||
    1123 CAATCTTATTTACCACAAGGAGATTCCTTTTTACAATCACTTGGACTTTA 1172

     390 leTrpGlyLeuAspAlaProHisArgMetTyrAsnGluIleIleHisLeu 406
         |||||::::::|||||||||:::::::::|||||:::|||:::   :::
    1173 TCTGGGCAATGGATGCCCCTCAAGAAGTTTACAATGACATTGTTTCTATG 1222

     407 MetGlnGlnGluGlu 411
         :::::::::::::::
    1223 ATATCAGAAGATAAA 1237

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH57420

seq_documentation_block:
ID   AAH57420 standard; cDNA; 1695 BP.
XX
AC   AAH57420;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human stomach cell specific cDNA sequence SEQ ID NO:260.
XX
```

```
KW   Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
KW   lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW   metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW   neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX
OS   Homo sapiens.
XX
PN   WO200132927-A2.
XX
PD   10-MAY-2001.
XX
PF   02-NOV-2000; 2000WO-US30396.
XX
PR   04-NOV-1999;   99US-0163508.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Sornasse T,  Seilhamer JJ,  Watson GA;
XX
DR   WPI; 2001-291057/30.
XX
PT   New cell and tissue specific polynucleotides useful for diagnosis,
PT   prognosis or monitoring of treatments for disorders where the gene is
PT   associated with a cancer, immunopathology or neuropathology -
XX
PS   Claim 1; Page 189-190; 327pp; English.
XX
CC   AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
CC   sequences (I). (I) can have cytostatic, immunomodulatory and
CC   neuroprotective activities, and can be used in gene therapy. (I) and
CC   proteins (II) encoded by then are used in high throughput screening
CC   assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC   mimetics, peptides, proteins, agonists, antagonists, antibodies or
CC   their fragments, immunoglobulins, inhibitors, drug compounds and
CC   pharmaceutical agents. Expression of (I) in a sample indicates the
CC   differentiation of embryonic stem cells into a tissue selected from
CC   brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC   tissues. (I) and (II) are used to produce an expression profile that
CC   defines a metabolic or developmental process, treatment, condition,
CC   disease or disorder. The gene profile can be used for diagnosis,
CC   prognosis or monitoring of treatments and for investigating a
CC   predisposition to a disorder where the gene is associated with a
CC   cancer, immunopathology or neuropathology.
XX
SQ   Sequence 1695 BP; 480 A; 356 C; 389 G; 470 T; 0 other;


alignment_scores:
           Quality: 1119.00              Length:    405
             Ratio:  3.401                Gaps:      2
 Percent Similarity:  81.235   Percent Identity: 51.852

alignment_block:
 US-09-578-063-47 x AAH57420   ..

 Align seg 1/1  to: AAH57420  from: 1  to: 1695

      14 ArgMetGluMetTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAs 30
         |||   :::|||||||||||||||:::   :::|||   ::::::
      31 AGGTCCAAAATGTGGCTGCTTTTAACAATGGCAAGTTTGATA........ 72

      30 nValAsnSerValHisMetProThrLysAlaVal.............. 41
                |||||          |||   ::::::
      73 .......TCTGTACTGGGGACTACACATGGTTTGTTTGGAAAATTACATC 115

      42 .....AspProGluAlaPheMetAsnIleSerGluIleIleGlnHisGln 56
              :::||||||:::   |||||||||||:::::::|||   :::
     116 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGG 165

      57 GlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIleLe 73
         |||||||||   |||||||||||||||:::||||||||||||||||||||
     166 GGATACCCAAATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCT 215

      73 uSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrGlyS 90
         |:::|||||||||||||||   |||       :::       :::|||||||:
     216 TGAAGTCAATAGAATTCCTTATGGGAAGAAAAATTCAGGGAATACAGGCC 265

      90 erArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsn 106
         ::|||||||||||||:::|||||||||||||||||:::::::::|||:::|||
     266 AGAGACCTGTTGTGTTTTTGCAGCATGGTTTGCTTGCATCAGCCACAAAC 315

     107 TrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAl 123
         ||||||||||||||||||||||||||||||:::||||||||||||||||||
     316 TGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTCTGGCAGATGC 365

     123 aGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgL 140
         ||||:::||||||||||:::||||||||||||||||||||:::|||:::|||:
     366 TGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 415

     140 ysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyr 156
         :::::          |||   |||:::   ||||||||||||||||||::
     416 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTT 465

     157 AspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLeuGl 173
         ||||||||||||:::::|||||||||||:::|||:::||||||:::::
     466 GATGAAATGGCTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAA 515

     173 nLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrT 190
         :|||:::|||||:::::::::::::|||||||||:::||||||||||||||
     516 GAAAGCTGGACAGAAGCAGCTACACTATGTTGGCCATTCCCAGGGCACCA 565

     190 hrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLysIle 206
         ||:::|||||||||||||||||||||   |||:::||||||:::::|||
     566 CCATTGGTTTTATTGCCTTTTCCACCAATCCCAGCCTGGCTAAAAGAATC 615

     207 LysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLysSe 223
         |||   ::::::|||||||||||||:::||||||||||||:::::::||||
     616 AAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGTATACAAAAAG 665

     223 rProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyLeuP 240
         |      :::|||:::   :::::::|||:::   :::::::|||   :::|
     666 CCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 715

     240 heGlyLysLysGluPheLeuTyrGlnThrArgPheLeuArgGlnLeuVal 256
         |||||   |||   |||       ::::::   |||   :::   |||:::
     716 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCT 765

     257 IleTyrLeuCysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMe 273
               :::|||::::::   :::|||:::   :::|||||||||   ::
     766 ACTGAAGTGTGCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTT 815

     273 tLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaS 290
         ::::::::::   ||||||:::::::::|||:::|||   ||||||   :
     816 ATTTATAATTTGTGGATTTGACAGTAAGAACTTTAACACGAGTCGCTTGG 865

     290 erValTyrAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeu 306
         ::||||||   :::|||:::   ||||||||||||||||||||::::::
     866 ATGTGTATCTATCACATAATCCAGCAGGAACTTCTGTTCAAAACATGTTC 915

     307 HisTrpSerGlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGl 323
         ||||||:::|||||||||:::||||||:::::::::|||:::|||||||||
     916 CATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTTATGACTGGGG 965

     323 ySerGluThrLysAsnLeuGluLysCysAsnGlnProThrProValArgT 340
         ||||   ::::::|||              :::|||      |||       |
     966 AAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1015

     340 yrArgValArgAspMetThrValProThrAlaMetTrpThrGlyGlyGln 356
         ||:::|||      |||:::||||||   |||:::|||:::||||||:::
    1016 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAG 1065

     357 AspTrpLeuSerAsnProGluAspValLysMetLeuLeuSerGluValTh 373
         |||   |||::::::|||:::||||||   :::||||||   ::::::
    1066 GACCTGTTGGCTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCC 1115

     373 rAsnLeuIleTyrHisLysAsnIleProGluTrpAlaHisValAspPheI 390
```

```
          ||||||||||||||||||:::||||||   :::   |||:::|||||||
    1116 CAATCTTATTTACCACAAGGAGATTCCTTTTTACAATCACTTGGACTTTA 1165

     390 leTrpGlyLeuAspAlaProHisArgMetTyrAsnGluIleIleHisLeu 406
         |||||:::::|||||||||:::::::::||||||:::|||:::   :::
    1166 TCTGGGCAATGGATGCCCCTCAAGAAGTTTACAATGACATTGTTTCTATG 1215

     407 MetGlnGlnGluGlu 411
         ::::::::::::::::
    1216 ATATCAGAAGATAAA 1230

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF28689

seq_documentation_block:
ID   AAF28689 standard; cDNA; 1308 BP.
XX
AC   AAF28689;
XX
DT   05-APR-2001  (first entry)
XX
DE   Human protein HP03372 coding sequence #2.
XX
KW   Human; hydrophobic domain; immune deficiency; autoimmune disorder;
KW   allergy; tissue growth; regeneration; wound healing; burn; tumour;
KW   periodontal disease; thrombolytic condition; haemostatic condition;
KW   infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102563-A2.
XX
PD   11-JAN-2001.
XX
PF   16-JUN-2000; 2000WO-JP03943.
XX
PR   02-JUL-1999;   99JP-0188835.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-071581/08.
DR   P-PSDB; AAB61608.
XX
PT   New human proteins with hydrophobic domains, useful for the treatment
PT   of immune disorders, tumors, allergic conditions, thrombosis and
PT   microbial infection -
XX
PS   Claim 4; Pages 122-124; 153pp; English.
XX
CC   The present invention relates to human proteins (AAB61608-AAB61617) and
CC   their coding sequences (AAF28679-AAF28698). The proteins of the present
CC   invention have hydrophobic domains and can be used for the treatment of
CC   various immune deficiencies and disorders, such as severe combined
CC   immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC   autoimmune pulmonary inflammation, graft-versus-host disease and
CC   Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC   of allergic reactions and conditions, such as asthma and in regulation of
CC   haematopoiesis or lymphoid cell deficiencies. The proteins may also have
CC   utility in compositions used for bone, cartilage, tendon and/or nerve
CC   tissue growth or regeneration as well as wound healing and in the
CC   treatment of burns. The proteins may be used in the treatment of
CC   periodontal disease and in other tooth repair processes. Other uses
CC   include treatment of thrombolytic and haemostatic conditions, treatment
CC   or prevention of tumours and inhibiting infection by bacteria, viruses,
CC   fungi and other parasites.
XX
SQ   Sequence 1308 BP; 402 A; 246 C; 264 G; 396 T; 0 other;


alignment_scores:
           Quality:  695.00            Length:    233
               Ratio:  3.717               Gaps:      2
 Percent Similarity:  80.258   Percent Identity: 56.223

alignment_block:
 US-09-578-063-47 x AAF28689   ..

 Align seg 1/1  to: AAF28689  from: 1  to: 1308

      14 ArgMetGluMetTrpLeuLeuIleLeuValAla..................... 24
         |||   :::|||||||   |||:::   :::|||
      67 AGATCCCAAATGTGGCAGCTTTTAGCAGCAGCATGCTGGATGCTTCTTCT 116

      25 ..........TyrMetPheGlnArgAsnValAsnSerValHisMetProT 38
                     |||   ::::::::::::   |||:::
     117 TGGATCTATGTATGGTTATGACAAGAAAGGAAACAAT.............. 153

      38 hrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGln 54
                 ::::::|||||||||   ||||||||||||:::|||||:::
     154 ........GCAAACCCTGAAGCTAATATGAATATTAGCCAGATTATTTCT 195

      55 HisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTy 71
         :::   |||||||||   ||||||||||:::|||:::|||:::||||||||
     196 TACTGGGGTTATCCTTATGAAGAGTATGATGTTACAACAAAAGATGGTTA 245

      71 rIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysT 88
         |||||||::::::   |||||||||:::|||          |||   :::|
     246 TATCCTTGGAATTTATAGGATTCCACATGGAAGAGGATGCCCAGGGAGGA 295

      88 hrGlySerArgProValValLeuGlnHisGlyLeuValGlyGlyAla 104
         ||:::   :::|||:::|||   |||||||||||||||:::::::::|||
     296 CAGCTCCAAAGCCTGCTGTGTATTTGCAGCATGGCTTAATTGCATCTGCC 345

     105 SerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAl 121
         ||||||||||||   |||||||||||||||||||||:::|||:::||||
     346 AGTAACTGGATTTGCAACCTGCCCAACAACAGTTTGGCTTTCCTTCTGGC 395

     121 aAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpS 138
         ||||:::|||:::||||||||:::|||||||||||||||||||:::||||
     396 AGATAGTGGTTATGACGTGTGGTTGGGGAACAGCCGAGGAAACACTTGGT 445

     138 erArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPhe 154
         ||||||||||      ||||||      :::   |||:::|||||||||
     446 CCAGAAAACACCTTAAATTGTCACCGAAATCACCAGAATACTGGGCCTTC 495

     155 SerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIl 171
         |||   ||||||||||||:::::|||||||||||:::||||||||||||
     496 AGTTTGGATGAGATGGCTAAATATGACCTTCCAGCCACAATCAATTTTAT 545

     171 eLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnG 188
         |:::::::||||||||||||:::::::::|||||||||||:::|||||||
     546 CATAGAGAAAACTGGACAGAAGCGACTCTACTACGTGGGCCACTCACAAG 595

     188 lyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGln 204
         ||||||||:::::|||||||||||||||||||   ||||||||||||:::
     596 GCACCACCATAGCTTTTATAGCATTTTCTACAAACCCAGAACTGGCTAAA 645

     205 LysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAl 221
         |||||||||:::::|||||||||||||||||:::::|||||||||:::::
     646 AAGATTAAGATATTTTTTGCACTGGCTCCAGTTGTCACAGTTAAATACAC 695

     221 aLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLys 237
         ::::||||||      |||:::      |||          ::::::|||
     696 CCAAAGTCCTATGAAAAAACTAACAACCCTTTCCAGGCGAGTAGTTAAG 744

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF28679

seq_documentation_block:
ID   AAF28679 standard; cDNA; 699 BP.
XX
AC   AAF28679;
XX
DT   05-APR-2001  (first entry)
XX
```

```
DE   Human protein HP03372 coding sequence #1.
XX
KW   Human; hydrophobic domain; immune deficiency; autoimmune disorder;
KW   allergy; tissue growth; regeneration; wound healing; burn; tumour;
KW   periodontal disease; thrombolytic condition; haemostatic condition;
KW   infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102563-A2.
XX
PD   11-JAN-2001.
XX
PF   16-JUN-2000; 2000WO-JP03943.
XX
PR   02-JUL-1999;   99JP-0188835.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-071581/08.
DR   P-PSDB; AAB61608.
XX
PT   New human proteins with hydrophobic domains, useful for the treatment
PT   of immune disorders, tumors, allergic conditions, thrombosis and
PT   microbial infection -
XX
PS   Claim 3; Page 113; 153pp; English.
XX
CC   The present invention relates to human proteins (AAB61608-AAB61617) and
CC   their coding sequences (AAF28679-AAF28698). The proteins of the present
CC   invention have hydrophobic domains and can be used for the treatment of
CC   various immune deficiencies and disorders, such as severe combined
CC   immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC   autoimmune pulmonary inflammation, graft-versus-host disease and
CC   Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC   of allergic reactions and conditions, such as asthma and in regulation of
CC   haematopoiesis or lymphoid cell deficiencies. The proteins may also have
CC   utility in compositions used for bone, cartilage, tendon and/or nerve
CC   tissue growth or regeneration as well as wound healing and in the
CC   treatment of burns. The proteins may be used in the treatment of
CC   periodontal disease and in other tooth repair processes. Other uses
CC   include treatment of thrombolytic and haemostatic conditions, treatment
CC   or prevention of tumours and inhibiting infection by bacteria, viruses,
CC   fungi and other parasites.
XX
SQ   Sequence 699 BP; 216 A; 144 C; 148 G; 191 T; 0 other;


alignment_scores:
            Quality:  691.00              Length:    221
              Ratio:  3.735                 Gaps:      2
 Percent Similarity:  83.710   Percent Identity: 58.371

alignment_block:
 US-09-578-063-47 x AAF28679   ..

 Align seg 1/1  to: AAF28679  from: 1  to: 699

      18 TrpLeuLeuIleLeuValAla...TyrMetPheGlnArgAsnValAsnSe 33
         |||:::|||:::|||   :::   |||   ::::::::::::   |||::
      28 TGGATGCTTCTTCTTGGATCTATGTATGGTTATGACAAGAAAGGAAACAA 77

      33 rValHisMetProThrLysAlaValAspProGluAlaPheMetAsnIleS 50
         :                    ::::::!!!!!!!!!   !!!!!!!!!!
      78 T.....................GCAAACCCTGAAGCTAATATGAATATTA 106

      50 erGluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAla 66
         ||:::!!!!!!::::::   !!!!!!!!   !!!!!!!!!:::|||:::
     107 GCCAGATTATTTCTTACTGGGGTTATCCTTATGAAGAGTATGATGTTACA 156

      67 ThrGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuVa 83
         |||:::!!!!!!!!!!!!!!!!!!::::::   !!!!!!!!!:::|||
     157 ACAAAAGATGGTTATATCCTTGGAATTTATAGGATTCCACATGGAAGAGG 206

      83 lGlnProLysLysThrGlySerArgProValValLeuLeuGlnHisGlyL 100
             |||   :::|||:::   :::|||:::!!!   !!!!!!!!!!!!
     207 ATGCCCAGGGAGGACAGCTCCAAAGCCTGCTGTGTATTTGCAGCATGGCT 256

     100 euValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeu 116
         ||:::::::::!!!!!!!!!!!!!!!!   !!!!!!!!!!!!!!!!!!!!
     257 TAATTGCATCTGCCAGTAACTGGATTTGCAACCTGCCCAACAACAGTTTG 306

     117 GlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSerAr 133
         :::|||:::!!!!!!!!!:::!!!:::!!!!!!!!!:::!!!!!!!!!!
     307 GCTTTCCTTCTGGCAGATAGTGGTTATGACGTGTGGTTGGGGAACAGCCG 356

     133 gGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAspG 150
         !!!!!!!:::!!!!!!!!!!!!!!!!      !!!!!!      :::   |
     357 AGGAAACACTTGGTCCAGAAAACACCTTAAATTGTCACCGAAATCACCAG 406

     150 luPheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeuProAla 166
         !!:::!!!!!!!!!!!!   !!!!!!!!!!!!:::::!!!!!!!!!!!!
     407 AATACTGGGCCTTCAGTTTGGATGAGATGGCTAAATATGACCTTCCAGCC 456

     167 ValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrVa 183
         :::!!!!!!!!!!!!!:::::!!!!!!!!!!!!:::::::::!!!!!!!!
     457 ACAATCAATTTTATCATAGAGAAAACTGGACAGAAGCGACTCTACTACGT 506

     183 lGlyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMetP 200
         !!!!:::!!!!!!!!!!!!!!!:::::!!!!!!!!!!!!!!!!!!   |
     507 GGGCCACTCACAAGGCACCACCATAGCTTTTATAGCATTTTCTACAAACC 556

     200 roGluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProIleAla 216
         !!!!!!!!!!!:::!!!!!!!!!:::::!!!!!!!!!!!!!!!!!:::::
     557 CAGAACTGGCTAAAAAGATTAAGATATTTTTTGCACTGGCTCCAGTTGTC 606

     217 ThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeuProAs 233
         !!!!!!!!!:::::::::!!!!!!      !!!:::      |||
     607 ACAGTTAAATACACCCAAAGTCCTATGAAAAAACTAACAACCCTTTCCAG 656

     233 pMetMetIleLys 237
             :::::::|||
     657 GCGAGTAGTTAAG 669

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA44349

seq_documentation_block:
ID   AAA44349 standard; cDNA; 617 BP.
XX
AC   AAA44349;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human secreted expressed sequence tag SEQ ID NO:924.
XX
KW   Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW   vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW   tumour; infection; depression; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200021991-A1.
```

```
XX
PD   20-APR-2000.
XX
PF   15-OCT-1999;   99WO-US24206.
XX
PR   15-OCT-1998;   98US-0104436.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR;
XX
DR   WPI; 2000-317938/27.
XX
PT   Isolated polynucleotides, and encoded proteins, comprising secreted
PT   expressed sequence tags (sESTs), useful for treating various disorders
PT   such as autoimmune, infectious, and central nervous system disorders -
XX
PS   Claim 1; Page 441; 803pp; English.
XX
CC   AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC   sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC   tissue sources. The sESTs can have a range of activities depending on
CC   the tissues they were isolated from. The activities include:
CC   chemotactic; proliferative; immunomodulatory; haematopoietic;
CC   chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC   cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC   antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC   nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC   anticonvulsant; and antidepressant. The sESTs can be used for gene
CC   therapy and in vaccines. The sESTs are useful as probes for the
CC   identification and isolation of full-length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporosis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC   disease), tumours, bacterial, fungal or viral infections, depression and
CC   psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 617 BP; 181 A; 125 C; 139 G; 172 T; 0 other;


alignment_scores:
           Quality:  596.00            Length:    190
             Ratio:  3.772               Gaps:      2
 Percent Similarity:  83.158   Percent Identity: 60.526

alignment_block:
 US-09-578-063-47 x AAA44349   ..

 Align seg 1/1  to: AAA44349  from: 1  to: 617

      16 GluMetTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAs 32
         :::|||||||||||||:::   :::|||   ::::::
      54 AAAATGTGGCTGCTTTTAACAATGGCAAGTTTGATA.............. 89

      32 nSerValHisMetProThrLysAlaVal.....................A 42
          ||||||         |||   ::::::                    :
      90 .TCTGTACTGGGGACTACACATGGTTTGTTTGGAAAATTACATCCTGGAA 138

      42 spProGluAlaPheMetAsnIleSerGluIleIleGlnHisGlnGlyTyr 58
         ::||||||:::   |||||||||||:::::|||   :::   ||||||
     139 GCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGGGGATAC 188

      59 ProCysGluGluTyrGluValAlaThrGluAspGlyTyrIleLeuSerVa 75
         |||   ||||||||||||||:::|||||||||||||||||||||:::||
     189 CCAAATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCTTGAAGT 238

      75 lAsnArgIleProArgGlyLeuValGlnProLysLysThrGlySerArgP 92
         ||||||||||||   |||      :::      :::|||||:::||||
     239 CAATAGAATTCCTTATGGGAAGAAAAATTCAGGGAATACAGGCCAGAGAC 288

      92 roValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsnTrpIle 108
         |||||||:::||||||||||||||||:::::::::|||:::||||||||
     289 CTGTTGTGTTTTTGCAGCATGGTTTGCTTGCATCAGCCACAAACTGGATT 338

     109 SerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAlaGlyPh 125
         ||||||||||||||||||||||||:::|||||||||||||||||||::
     339 TCCAACCTGCCGAACAACAGCCTTGCCTTCATTCTGGCAGATGCTGGTTA 388

     125 eAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgLysHisL 142
         :|||||||||:::|||||||||||||||||||:::|||:::|||::::::
     389 TGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAAGAAACT 438

     142 ysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyrAspGlu 158
                 |||   |||:::   ||||||||||||||||||:::||||||
     439 TGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAA 488

     159 MetAlaArgPheAspLeuProAlaValIleAsnPheIleLeuGlnLysTh 175
         ||||||::::::||||||||||||:::|||:::||||||::::::|||||
     489 ATGGCTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAAGAAAAC 538

     175 rGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMetG 192
         |||||||:::::::::::::|||||||||:::|||||||||||||||:::|
     539 TGGACAGAAGCAGCTACACTATGTTGGCCATTCCCAGGGCACCACCATTG 588

     192 lyPheIleAlaPheSerThr 198
         ||||||||||||||||||
     589 GTTTTATTGCCTTTTCCACC 608
```

OM of: US-09-578-063-47 to: Issued_Patents_NA:* out_format : pfs

Date: Jan 31, 2002 12:05 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09578063/runat_30012002_092151_26325/app_query.fasta_1.488
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09578063_@CGN1_1_81 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-578-063-47
Query length: 423
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 90.070000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-2	+	1141.50	2461.53	1.9e-129	1137	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-2	+	1141.50	2461.53	1.9e-129	1137	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-4	+	1141.50	2461.49	1.9e-129	1140	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-4	+	1141.50	2461.49	1.9e-129	1140	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-6	+	1141.50	2461.43	1.9e-129	1146	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-6	+	1141.50	2461.43	1.9e-129	1146	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-1	+	1141.50	2457.71	3.0e-129	1528	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-1	+	1141.50	2457.71	3.0e-129	1528	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-356-952-12	+	271.50	510.09	9.2e-21	43676	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-751-782-2	+	155.00	326.10	1.6e-10	178	! S
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-925-171-2	+	155.00	326.10	1.6e-10	178	! S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-484-105-21	+	105.00	187.71	0.0083	1676	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-484-106-21	+	105.00	187.71	0.0083	1676	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-602-359A-24	+	102.00	187.29	0.0088	1041	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-232-519-1	+	98.50	178.16	0.0284	1167	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-956-1	+	98.50	178.16	0.0284	1167	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-115-446-1	+	92.00	153.99	0.6298	2521	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-115-446-5	+	92.00	153.96	0.6318	2526	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-119A-1	+	92.00	116.26	79.56	46899	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-080-386-3	-	91.00	153.79	0.6457	2160	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-390-000A-3	-	91.00	153.79	0.6457	2160	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-742-185-102	-	91.00	116.02	82.06	40328	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-968-542C-1	+	89.00	136.17	6.19	6027	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-755-559-2	-	85.00	140.54	3.53	2180	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-210-474-2	-	85.00	140.54	3.53	2180	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-602-359A-30	+	84.50	149.29	1.15	1017	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-759-945-1	+	84.50	137.05	5.53	2625	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-707-367-1	+	84.50	133.77	8.41	3383	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-386-607-1	+	84.00	138.47	4.61	2160	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-882-329-4	+	82.50	141.46	3.14	1329	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-183-213-4	+	82.50	141.46	3.14	1329	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-446-803-4	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-861-837-4	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-600-656-4	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-170-670-9	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-170-670-14	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-193-068-9	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-193-068-13	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-183-412-9	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-183-412-13	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-354-191A-4	+	82.50	140.29	3.65	1455	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-600-908A-11	+	82.50	140.26	3.66	1458	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-683-838A-11	+	82.50	140.26	3.66	1458	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-483-101-1	+	82.50	122.44	35.99	5798	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-348-891A-1	+	82.50	109.43	191.04	15894	!

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-2

seq_documentation_block:
; Sequence 2, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
US-08-227-108-2

alignment_scores:
Quality: 1141.50 Length: 372
Ratio: 3.624 Gaps: 3
Percent Similarity: 84.677 Percent Identity: 56.452

alignment_block:
US-09-578-063-47 x US-08-227-108-2 ..

Align seg 1/1 to: US-08-227-108-2 from: 1 to: 1137

```
      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
         ||||||         :::|||||:::   |||||||||||:::::||
      19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
         |   :::   |||||||||:::||||||||||||||||:::||||||||||
      60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

      70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
```

```
     |||||||||||:::::::::|||||||||   |||      :::   :::
 110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

  87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
     :::   |||   |||||||||:::::::||||||||||||||||:::::::
 160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

 103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
     :|||:::|||||||||||||||||||||||||||||||:::|||||||
 210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

 120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
     ||||||||||||||:::|||||||||:::|||||||||||||||||||:::
 260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

 137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
     |||:::|||::::::          |||   |||:::   ||||||||||
 310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 359

 153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
     |||||||:::|||||||||||:::::|||||||||||:::|||:::|
 360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409

 170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
     ||||||||:::|||||||||||:::|||:::::|||||||||:::|||
 410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

 187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
     ||||||||||:::|||||||||||||||||||   |||:::|||||
 460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

 203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
     |:::::|||||   ::::::|||||||||||:::|||||||||||:
 510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

 220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
     ::::::::::::      :::|||::::::|||:::|||:::::::::::
 560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

 237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
     |||   :::||||||:::|||   |||   |||       :::|||:::
 610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

 253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
      |||   |||:::      :::|||::::::      :::|||   :::|
 657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

 269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
     ||||||||   ::::::::::::   ||||||:::|||   |||:::|||
 707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

 286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
     ||||||||   :::||||||   :::|||:::   ||||||||||||||
 757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

 302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
     |||||||:::|||||||||||||||||||||:::||||||:::::::::|
 807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

 319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
     ||||||||||||||||   :::::::|||:::          :::|||
 857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

 336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
        |||      |||::::::   ||||||   ||||||   |||:::||
 907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

 352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
     |:::||||||:::|||   |||:::::::|||:::||||||   :::|||||
 957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

 369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
     |||||::::::   ||||||||||||||||:::::||||||   :::
```

```
    1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

     386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
         |||:::|||||||||||||:::::::||||||||||:::   :::||||||||
    1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

     402 uIleIleHisLeuMet 407
         ||||:::   :::|||
    1107 AATTGTTTCCATGATG 1122

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-2

seq_documentation_block:
; Sequence 2, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1137 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1137
US-09-073-674-2


alignment_scores:
            Quality: 1141.50              Length:    372
              Ratio:  3.624                Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
 US-09-578-063-47 x US-09-073-674-2   ..

 Align seg 1/1  to: US-09-073-674-2  from: 1  to: 1137

      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
         ||||||          :::|||||||:::   |||||||||||:::::||
      19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
```

```
    |   :::   |||||||||:::|||||||||||||||||:::||||||||||
 60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

 70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
    ||||||||||:::::::::|||||||||   |||      :::   :::
110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

 87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
    :::   |||   |||||||||:::::|||||||||||||||||:::::::
160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
    :|||:::|||||||||||||||||||||||||||||||:::|||||||
210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
    ||||||||||||||:::|||||||||:::||||||||||||||||||:::
260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
    |||:::|||:::::          |||   |||:::   ||||||||||
310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 359

153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
    |||||||:::|||||||||||||:::::|||||||||||||:::|||:::|
360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409

170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
    ||||||||:::|||||||||||||:::|||:::::|||||||||:::|||
410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
    ||||||||||||:::|||||||||||||||||||||||   |||:::|||||
460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
    |:::::|||||   :::::|||||||||||||:::|||||||||||||:
510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
    ::::::::::      :::|||:::::::|||:::|||:::::::::::::
560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
    |||   :::||||||:::|||   |||   |||      :::|||:::
610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
     |||   |||:::      :::|||::::::      :::|||   :::|
657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
    ||||||||   ::::::::::::   ||||||:::|||   |||:::|||
707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
    |||||||||   :::|||||||   :::|||:::   ||||||||||||||
757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
    |||||||:::|||||||||||||||||||||:::|||||:::::::::|
807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
    ||||||||||||||||   :::::::|||:::         :::|||
857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
       |||      |||::::::   ||||||   ||||||   |||:::||
907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
    |:::|||||:::|||   |||::::::|||:::||||||   :::|||||
```

```
957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

 369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
     |||||:::::::   |||||||||||||||:::::||||||    :::
1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

 386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
     |||:::||||||||||||:::::||||||||||:::   :::|||||||||
1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

 402 uIleIleHisLeuMet 407
     ||||:::   :::|||
1107 AATTGTTTCCATGATG 1122
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-4

```
seq_documentation_block:
; Sequence 4, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1140
US-08-227-108-4


alignment_scores:
            Quality: 1141.50              Length:    372
              Ratio:  3.624                Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
 US-09-578-063-47 x US-08-227-108-4   ..

 Align seg 1/1  to: US-08-227-108-4  from: 1  to: 1140

      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
```

```
 22 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 62

 53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
 63 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 112

 70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
113 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 162

 87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
163 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 212

103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
213 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 262

120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
263 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 312

137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
313 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 362

153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
363 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 412

170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
413 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 462

187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
463 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 512

203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
513 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 562

220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
563 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 612

237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
613 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 659

253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
660 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 709

269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
710 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 759

286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrServa 302
760 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 809

302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
810 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 859

319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
860 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 909

336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
```

```
910 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 959

352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
960 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1009

369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
1010 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1059

386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
1060 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1109

402 uIleIleHisLeuMet 407
1110 AATTGTTTCCATGATG 1125
```

```
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-4

seq_documentation_block:
; Sequence 4, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1140
US-09-073-674-4


alignment_scores:
            Quality: 1141.50              Length:    372
              Ratio:  3.624                Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
 US-09-578-063-47 x US-09-073-674-4    ..
```

```
Align seg 1/1  to: US-09-073-674-4  from: 1  to: 1140

     37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
     22 CCCACA.....AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 62

     53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
     63 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 112

     70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
    113 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 162

     87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
    163 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 212

    103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
    213 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 262

    120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
    263 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 312

    137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
    313 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 362

    153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
    363 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 412

    170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
    413 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 462

    187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
    463 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 512

    203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
    513 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 562

    220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
    563 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 612

    237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
    613 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 659

    253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
    660 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 709

    269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
    710 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 759

    286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
    760 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 809

    302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
    810 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 859

    319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335

    860 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 909

    336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
    910 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 959

    352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
    960 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1009

    369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
   1010 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1059

    386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
   1060 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1109

    402 uIleIleHisLeuMet 407
   1110 AATTGTTTCCATGATG 1125
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-6

```
seq_documentation_block:
; Sequence 6, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1146 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-227-108-6


alignment_scores:
            Quality: 1141.50              Length:    372
              Ratio:  3.624                Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
```

```
US-09-578-063-47 x US-08-227-108-6   ..

Align seg 1/1  to: US-08-227-108-6  from: 1  to: 1146

     37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
        ||||||          :::|||||||:::   ||||||||||||:::::||
     19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

     53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
        |   :::   |||||||||:::||||||||||||||||:::||||||||||
     60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

     70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
        ||||||||||:::::::::|||||||||   |||      :::   :::
    110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

     87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
        :::   |||   |||||||||:::::|||||||||||||||||:::::::
    160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

    103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
        :|||:::|||||||||||||||||||||||||||||||:::|||||||
    210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

    120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
        ||||||||||||||:::|||||||||:::|||||||||||||||||:::
    260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

    137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
        |||:::|||::::::          |||   |||:::   ||||||||||
    310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 359

    153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
        |||||||:::|||||||||||||:::::|||||||||||||:::|||:::|
    360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409

    170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
        ||||||||:::|||||||||||||:::|||:::::::|||||||||:::|||
    410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

    187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
        ||||||||||||:::|||||||||||||||||||||   |||:::|||||
    460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

    203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
        |:::::|||||   :::::|||||||||||||:::|||||||||||||:
    510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

    220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
        ::::::::::      :::|||:::::::|||:::|||:::::::::::
    560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

    237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
        |||   :::||||||:::|||   |||   |||      :::|||:::
    610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

    253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
         |||   |||:::      :::|||:::::      :::|||   :::|
    657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

    269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
        ||||||||   ::::::::::::   ||||||:::|||   |||:::|||
    707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

    286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
        |||||||||   :::|||||||   :::|||:::   ||||||||||||||
    757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

    302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
        |||||||:::|||||||||||||||||||||||:::||||||:::::::::|
    807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

    319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335

        ||||||||||||||||   ::::::|||:::         :::|||
    857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

    336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
           |||      |||::::::   ||||||   ||||||   |||:::||
    907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

    352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
        |:::|||||:::|||   |||:::::|||:::|||||   :::||||
    957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

    369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
        |||||::::::   ||||||||||||||||:::::||||||   :::
   1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

    386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
        |||:::||||||||||||:::::||||||||||:::   :::|||||||||
   1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

    402 uIleIleHisLeuMet 407
        ||||:::   :::|||
   1107 AATTGTTTCCATGATG 1122

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-6

seq_documentation_block:
; Sequence 6, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1146 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-09-073-674-6


alignment_scores:
            Quality: 1141.50              Length:    372
              Ratio:  3.624                Gaps:      3
 Percent Similarity:  84.677   Percent Identity: 56.452
```

```
alignment_block:
 US-09-578-063-47 x US-09-073-674-6   ..

 Align seg 1/1  to: US-09-073-674-6  from: 1  to: 1146

      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
         ||||||         :::|||||||:::   ||||||||||||:::::||
      19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
         |   :::   |||||||||:::|||||||||||||||:::||||||||||
      60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

      70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
         ||||||||||:::::::::|||||||||   |||       :::   :::
     110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

      87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
         :::   |||   |||||||||:::::|||||||||||||||||:::::::
     160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

     103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
         :|||:::||||||||||||||||||||||||||||||||:::|||||||
     210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

     120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
         ||||||||||||:::||||||||||::||||||||||||||||||:::
     260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

     137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
         |||:::|||::::::        |||   |||:::   ||||||||||
     310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 359

     153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
         |||||||:::|||||||||||||:::::|||||||||||||:::|||:::|
     360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409

     170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
         ||||||||:::|||||||||||:::|||:::::::||||||||:::|||
     410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

     187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
         ||||||||||||:::|||||||||||||||||||||   |||:::|||||
     460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

     203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
         |:::::::||||||   ::::::|||||||||||||:::|||||||||||:
     510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

     220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
         ::::::::::      :::|||:::::::|||:::|||::::::::::
     560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

     237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
         |||   :::|||||||:::|||   |||   |||       :::|||:::
     610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

     253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
         |||   |||:::      :::|||::::::      :::|||   :::|
     657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

     269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
         ||||||||   ::::::::::::   ||||||:::|||   |||:::|||
     707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

     286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
         |||||||||   :::||||||   :::|||:::   ||||||||||||||
     757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

     302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
         |||||||:::|||||||||||||||||||||:::||||||:::::::::|
     807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

     319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
         ||||||||||||||||   ::::::|||:::         :::|||
     857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

     336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
         |||      |||::::::   ||||||   ||||||   |||:::||
     907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

     352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
         |:::||||||:::|||   |||::::::|||:::||||||   :::||||
     957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

     369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
         |||||::::::   ||||||||||||||:::::||||||||   :::
    1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

     386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
         |||:::||||||||||||:::::||||||||||:::   :::||||||||
    1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

     402 uIleIleHisLeuMet 407
         ||||:::   :::|||
    1107 AATTGTTTCCATGATG 1122

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-227-108-1

seq_documentation_block:
; Sequence 1, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1528 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-227-108-1


alignment_scores:
            Quality: 1141.50              Length:    372
              Ratio: 3.624                  Gaps:      3
 Percent Similarity: 84.677    Percent Identity: 56.452
```

```
alignment_block:
 US-09-578-063-47 x US-08-227-108-1   ..

 Align seg 1/1  to: US-08-227-108-1  from: 1  to: 1528

      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
         ||||||         :::|||||||:::   |||||||||||||:::::||
      19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
         |   :::   ||||||||||:::|||||||||||||||||:::|||||||||||
      60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

      70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
         ||||||||||:::::::::|||||||||  |||      :::   :::
     110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

      87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
         :::   |||   |||||||||:::::::|||||||||||||||:::::::
     160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

     103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
         :|||:::|||||||||||||||||||||||||||||||||||:::|||||||
     210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

     120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
         ||||||||||||||:::|||||||||:::|||||||||||||||||||:::
     260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

     137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
         |||:::|||:::::          |||   |||:::   ||||||||||
     310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 359

     153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
         |||||||:::|||||||||||||:::::||||||||||||:::|||:::|
     360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409

     170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
         ||||||||:::|||||||||||||:::|||:::::||||||||||:::|||
     410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

     187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
         ||||||||||||:::|||||||||||||||||||||   |||:::||||||
     460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

     203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
         |:::::||||||   ::::::|||||||||||||:::||||||||||||:
     510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

     220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
         ::::::::::      :::|||:::::::|||:::|||:::::::::::
     560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

     237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
         |||   :::||||||:::|||   |||   |||      :::|||:::
     610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

     253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
          |||   |||:::      :::|||::::::      :::|||   :::|
     657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

     269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
         |||||||   ::::::::::::   |||||||:::|||   |||:::|||
     707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

     286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrServa 302
         |||||||||   :::||||||   :::|||:::   ||||||||||||||
     757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

     302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319
         |||||||:::|||||||||||||||||||||:::||||||:::::::::|
     807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

     319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
         ||||||||||||||||   ::::::|||:::         :::|||
     857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

     336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
            |||      |||::::::   ||||||   ||||||   |||:::||
     907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

     352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
         |:::|||||||:::|||   |||::::::|||:::||||||   :::||||
     957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

     369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
         |||||::::::   ||||||||||||||:::::::||||||   :::
    1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

     386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
         |||:::|||||||||||||:::::|||||||||:::   :::||||||||
    1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

     402 uIleIleHisLeuMet 407
         ||||:::   :::|||
    1107 AATTGTTTCCATGATG 1122

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-1

seq_documentation_block:
; Sequence 1, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0,.Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1528 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-09-073-674-1


alignment_scores:
            Quality: 1141.50              Length:    372
              Ratio:  3.624                Gaps:      3
```

```
 Percent Similarity:  84.677   Percent Identity: 56.452

alignment_block:
 US-09-578-063-47 x US-09-073-674-1   ..

 Align seg 1/1  to: US-09-073-674-1  from: 1  to: 1528

      37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIl 53
      19 CCCACA.........AACCCTGAAGTGACCATGAATATAAGTCAGATGAT 59

      53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspG 70
      60 CACCTACTGGGGATACCCAGCTGAGGAATATGAAGTTGTGACCGAAGACG 109

      70 lyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
     110 GTTATATCCTTGGGATCGACAGAATTCCTTATGGGAGGAAAAATTCAGAG 159

      87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyGl 103
     160 AATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCGCATC 209

     103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120
     210 AGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCC 259

     120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
     260 TGGCCGACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACC 309

     137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
     310 TGGGCCAGGAGGAATCTGTACTACTCGCCCGACTCCGTCGAATTCTGGGC 359

     153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
     360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTTCCCGCCACCATTGACT 409

     170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
     410 TCATCTTGAAGAAAACGGGACAGGACAAGCTACACTACGTTGGCCATTCC 459

     187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
     460 CAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGCTGGC 509

     203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
     510 GAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGT 559

     220 isAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
     560 ACACCGAAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTC 609

     237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
     610 AAGCTTATATTTGGAAACAAAATATTC...TACCCACACCACTTCTTTGA 656

     253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
     657 TCAATTTCTCGCCACCGAGGTATGCTCCCGCGAGACGGTGGATCTCCTCT 706

     269 ysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
     707 GCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTATGAACTTGAAC 756

     286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
     757 ATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGGT 806

     302 lGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgA 319

     807 TCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAG 856

     319 laPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
     857 CTTTTGACTGGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGC 906

     336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
     907 ATGCCTCCCTACTACAACCTGACAGACATGCATGTGCCAATCGCAGTGTG 956

     352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
     957 GAACGGTGGCAACGACTTGCTGGCCGACCCTCACGATGTTGACCTTTTGC 1006

     369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
    1007 TTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTCCTCCTTACAAT 1056

     386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGl 402
    1057 CACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATGA 1106

     402 uIleIleHisLeuMet 407
    1107 AATTGTTTCCATGATG 1122

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-356-952-12

seq_documentation_block:
; Sequence 12, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
;  APPLICANT: Boriack-Sjodin, Ann
;  APPLICANT: Margarit, S. M.
;  APPLICANT: Bor-Sogi, Dafna
;  APPLICANT: Cole, Philip
;  APPLICANT: Kuriyan, John
;  TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: 600-1-228N
;  CURRENT APPLICATION NUMBER: US/09/356,952
;  CURRENT FILING DATE: 1999-07-19
;  EARLIER APPLICATION NUMBER: 60/093,631
;  EARLIER FILING DATE: 1998-07-21
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 43676
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-356-952-12


alignment_scores:
           Quality:  271.50            Length:    376
             Ratio:  1.263              Gaps:     16
 Percent Similarity:  57.181   Percent Identity: 25.532

alignment_block:
 US-09-578-063-47 x US-09-356-952-12   ..

 Align seg 1/1  to: US-09-356-952-12  from: 1  to: 43676

      61 GluGluTyrGluValAlaThrGluAspGlyTyrIleLeuSerVal...As 76
   14769 GAGGAATTTAGATTGGAGACCGAAGATGGGTTTGTTATAGATTTGTGGCA 14818

      76 nArgIleProArgGlyLeuValGlnProLysLysThrGlySerArgProV 93
   14819 CTTGATACCAAAATATAGAACGACAGATTCTGACAAGAAGAAGAGGCCAC 14868

      93 alValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsnTrpIleSer 109
```

```
14869 CCATTTTGATGCTACATGGCCTTTTGCAAAGCAGTGGTTCGTTCGCATCC 14918

  110 AsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAlaGlyPheAs 126
      |||      ::::::||||||::::::::::|||   :::::::|||:::||
14919 AAT...GGTAGAAAATCTCTGGCATATTTTCTGTATCAATCCGGTTACGA 14965

  126 pValTrpMetGlyAsnSerArgGlyAsn.........AlaTrpSerArg. 139
      |:::|||:::||||||:::|||   :::             |||:::::::
14966 CATATGGTTAGGGAATAACAGATGCGGGTTTAGGCCGGAATGGAACGAAG 15015

  140 ..LysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
        |||      ||||||:::             :::   |||   ::::::
15016 CGAAAGTACCGACACTAGCT............TCCAGGTGGGACTGGGAC 15053

  156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
            ||||||::::::::::||||||      :::|||:::   :::||
15054 CTTCGCGAGATGGTTAAGTACGATCTGACCCTTTTGATTGATACCGTGTT 15103

  172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
      |   ||||||      ||||||:::      :::::::::||||||||||
15104 AGCTAAGACGCAGTTTGAAAAGCTTACTTTGATCTCGCATTCTCAGGGCA 15153

  189 hrThrMetGlyPheIleAlaPhe..................... 196
      |||||:::||||||:::::::::
15154 CTACACAGGGGTTTATGGGCTTGGTCAACGAAGATAAGTTTTTCCCTCCC 15203

  197 .........SerThrMetProGluLeuAlaGlnLysIleLysMetTyrPh 210
              |||             ::::::::::||||||      |||::
15204 GGTTCGGGATCTAAAGAATCTTTTTTCACTTCTAAGATCGCAAACTATAT 15253

  210 eAlaLeuAlaProIleAlaThrValLysHisAlaLysSerProGlyThrL 227
      :|||||||||||                |||      ||||||
15254 TGCCTTGGCCCCC.................GCAGTGTATCCTGGTCCC. 15284

  227 ysPheLeuLeuLeuProAspMetMetIleLys.................. 237
            ||||||         ::::::::::|||
15285 .....TTACTTAACGAGAAATTGTTTGTTAAGCTTATGACAAAGGAAATC 15329

  238 .........GlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPh 251
                 :::|||||||:::   :::|||:::             ::
15330 GAAAATCCCTGGTTCTTTGGTGAAACGAGCTTTTTTGAGATAATGATGAT 15379

  251 eLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnI 268
      ::::|||:::|||   :::          |||:::   :::::::::   :
15380 TGTAAGAAACTTGTGCGTT.........GGTGAGAGCTTGTTCTCCTTTG 15420

  268 leCysSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMet 284
      ::|||   :::|||:::      |||      :::|||   :::
15421 TTTGTTACACCATCTTCAATTACCTGTTTGATTGGAACGATACCCTTTGG 15470

  285 AsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSe 301
      :::   :::      :::   :::          :::   :::   :::||
15471 GATACCGCATTAAGAGATCGCCATTTCCTGTTTTCGCCAGTCCATGTTTC 15520

  301 rValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuA 318
      ||||:::   :::::::::|||.     :::   :::   :::::::::
15521 AGTGAAG...TTGATGCAATGGTGGCTGTCACCCGACCCCAACAAGGTA. 15566

  318 rgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGln 334
        :::|||   :::|||||||:::
15567 ..AGTTTTAAATTTGGTTCCCATAAGATG.................TTC 15596

  335 ProThrProValArgTyrArgValArgAspMetThrValProThrAla.. 350
      |||      |||::::::                 :::|||:::
15597 CCCGACAATGTTAAGTGGTTTTCAGACGCATCAAAGGCCCCAAATATCTA 15646

  351 .MetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysM 367
       :::::::::      ||||||   |||            |||   :::
15647 CTTGTTTGTTCCAAAGCAAGATAGATTG.........GTGGACGGAGAAA 15687

  367 etLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsn......... 380
        |||::::::::::   :::|||:::      :::   |||
15688 GACTAATCAATCATTTCGTCAATGTGGAGTCGAATGTCAACTACAAGATC 15737
```

```
      381 ......IleProGluTrpAlaHisValAspPheIleTrpGlyLeuAspAl 395
                |||   |||:::||||||:::|||   :::|||:::   |||::
   15738 TGGTACATTGATGAGTATGCCCATATTGATGTCCTATGGGCACATGATGT 15787

      395 aProHisArgMetTyrAsnGluIleIle 404
          :   :::|||:::   :::   |||:::
   15788 CATAGAGAGAATTGGTAAACCAATTTTA 15815

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-751-782-2

seq_documentation_block:
; Sequence 2, Application US/08751782
; Patent No. 5821352
;  GENERAL INFORMATION:
;    APPLICANT:  Heintz, Nathaniel
;    APPLICANT:  Gubbay, Johnathan
;    APPLICANT:  Skinner, Michael
;    TITLE OF INVENTION:  A cDNA Library Prepared during
;    TITLE OF INVENTION:  Regression of Rat Prostate and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,782
;      FILING DATE:  18-NOV-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  178 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  10.2
US-08-751-782-2


alignment_scores:
            Quality:  155.00            Length:     50
              Ratio:  3.780              Gaps:      0
 Percent Similarity:  82.000   Percent Identity: 56.000

alignment_block:
 US-09-578-063-47 x US-08-751-782-2  ..

 Align seg 1/1  to: US-08-751-782-2  from: 1  to: 178

     139 ArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSe 155
         |||||||||      |||:::   |||:::   |||||||||   |||||
       7 AGGAAACATGTGCGCCTAGACCCAGATTCTAAAGAATTTTGGGATTTTAG 56

     155 rTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleL 172
```

```
          |:::::::|||:::   ::::::||||||||||||||:::|||   |||||||
       57 TTTTAATGAACAAATAGAATACGACCTCCCAGCCATCATTTATTTCATTC 106

      172 euGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGly 188
          ||:::::::|||   |||   :::|||||||||:::|||:::|||||||||
      107 TGAATGAAACAAGACAAACACAAATTTACTATATTGGCCATTCCCAGGGC 156

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-925-171-2

seq_documentation_block:
; Sequence 2, Application US/08925171
; Patent No. 5928871
;  GENERAL INFORMATION:
;    APPLICANT:  Heintz, Nathaniel
;    APPLICANT:  Gubbay, Johnathan
;    APPLICANT:  Skinner, Michael
;    TITLE OF INVENTION:  A cDNA Library Prepared during
;    TITLE OF INVENTION:  Regression of Rat Prostate and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/925,171
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-190 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  178 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  10.2
US-08-925-171-2


alignment_scores:
            Quality:  155.00            Length:     50
              Ratio:  3.780              Gaps:      0
 Percent Similarity:  82.000   Percent Identity: 56.000

alignment_block:
 US-09-578-063-47 x US-08-925-171-2   ..

 Align seg 1/1  to: US-08-925-171-2  from: 1  to: 178

      139 ArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSe 155
          |||||||||      |||:::   |||:::   |||||||||   |||||
        7 AGGAAACATGTGCGCCTAGACCCAGATTCTAAAGAATTTTGGGATTTTAG 56

      155 rTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleL 172
          |:::::::|||:::   ::::::||||||||||||||:::|||   |||||||
       57 TTTTAATGAACAAATAGAATACGACCTCCCAGCCATCATTTATTTCATTC 106

      172 euGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGly 188
          ||:::::::|||   |||   :::|||||||||:::|||:::|||||||||
      107 TGAATGAAACAAGACAAACACAAATTTACTATATTGGCCATTCCCAGGGC 156

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-484-105-21

seq_documentation_block:
; Sequence 21, Application US/08484105
; Patent No. 5589341
;  GENERAL INFORMATION:
;    APPLICANT:  STILLMAN, Bruce
;    APPLICANT:  BELL, Stephen P
;    APPLICANT:  KOBAYASHI, Ryuji
;    APPLICANT:  RINE, Jasper
;    APPLICANT:  FOSS, Margit
;    APPLICANT:  McNALLY, Francis J
;    APPLICANT:  LAURENSON, Patricia
;    APPLICANT:  HERSKOWITZ, Ira
;    APPLICANT:  LI, Joachim J
;    APPLICANT:  GAVIN, Kimberly
;    TITLE OF INVENTION:  ORIGIN OF REPLICATION COMPLEX GENES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,105
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59032/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1676 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  13..1302
US-08-484-105-21


alignment_scores:
            Quality:  105.00            Length:    278
              Ratio:  0.802              Gaps:     13
 Percent Similarity:  47.122   Percent Identity: 21.942

alignment_block:
 US-09-578-063-47 x US-08-484-105-21   ..

 Align seg 1/1  to: US-08-484-105-21  from: 1  to: 1676

       85 ProLysLysThrGlySerArgProValValLeuLeuGlnHisGlyLeuVa 101
          |||:::|||   |||||||||                 :::      ::
       85 CCAGAAAAAGAAGGATCCAGA.................CAGAAAAAGAC 116
```

```
    101 lGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsn..... 114
        ::::|||   :::|||         ||||||   :::|||
    117 GAATGGAAAAGAGAATGCTTCTAGAAATTTGCAATCAAATTTAGAAGAAG 166

    115 ........SerLeuGlyPhe..............IleLeuAlaAspAla 123
                ::::|||||||||                  :::|||::::::
    167 ATTTGGAACAACTGGGCTTCGAGGATGAAACTGTATCAATGGCTCAATCA 216

    124 GlyPheAspValTrp......MetGlyAsnSer............ArgGl 134
        :::::::::   :::      :::|||:::|||              ::
    217 GCAATCGAAAATTACTTTATGCAAGGAAAATCGGCGTCAGAACGAATGAA 266

    134 yAsnAlaTrpSerArgLysHisLys.................ThrLeuS 145
        :||||||   ||||||:::   :::                  |||   :
    267 TAATGCGAAATCCCGTCGTGGAAGACGTGCTGGAAATGGAAATACTGAAG 316

    145 erIleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMetAlaArg 161
        ::|||:::::|||:::      :::   :::      ::::::::::::
    317 AAATTGAGGAAGACGATGAGATCAGTAATGCTATCACTGATTTCACAAAA 366

    162 PheAspLeuProAlaValIleAsnPheIleLeuGlnLysThrGlyGlnGl 178
          |||||||||:::::   |||:::|||   :::|||   :::   ||
    367 TGTGATCTCCCTGGACTTCGAAATTATATTACCAAAAAAGATAACACGGA 416

    178 uLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMetGlyPheIleA 195
        |          :::   :::          :::   :::|||
    417 ATTCGAAAAACGATTGGAGCATCTCGCGGATAATGATTTCGGA..... 459

    195 laPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyrPheAla 211
                                    |||   |||:::|||:::|||
    460 ............................AAATGGAAGCTTTACCTAGCA 480

    212 LeuAlaProIleAlaThrValLysHisAlaLysSerProGlyThrLysPh 228
        :::
    481 GCTGGA..................................... 486

    228 eLeuLeuLeuProAspMetMetIleLysGlyLeuPheGlyLysLysGluP 245
                  ::::::::::::   |||:::   :::|||::::::
    487 ..........TTTAATATTCTTTTGCACGGTGTCGGTTCGAAGCGTGATG 526

    245 heLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTyrLeuCysGly 261
          |||      |||:::|||   :::::::|||      |||
    527 TTCTC......ACAGAATTTGAGAATGAGCTATCCGATTATACATATATG 570

    262 GlnValIleLeuAspGlnIleCysSerAsnIleMetLeuLeuLeuGlyGl 278
        :::|||         :::         |||:::   :::||||||||::
    571 AGAGTGGATGCACGGAAAGATGGGCTCAATGTAAAAGTTCTTCTTGGAGC 620

    278 yPheAsnThrAsn.........AsnMetAsnMetSerArgAlaSerValT 292
        ::::|||   |||         |||   |||::::::|||:::
    621 TATCAATGAGAATATGAAGCTGAATTGTAATGTGAAGAGAGGC..... 663

    292 yrAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrp 308
                                      |||:::   :::   |||
    664 ..............................CAATCTACGATTAGTTGG 681

    309 SerGlnAlaVal............AsnSerGlyGluLeu........... 317
        ::::::::::::            ||||||   :::|||
    682 GCTCGATCTATTCGCAGAAAAATGAATAGCCAACAGTTGATTCTTATCAT 731

    318 ..........ArgAlaPheAspTrpGlySerGlu 325
                ::::|||   ||||||   |||:::
    732 TGATAATATTGAAGCTCCTGATTGGAGAAGTGAT 765

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-484-106-21

seq_documentation_block:
; Sequence 21, Application US/08484106
; Patent No. 5614618
;  GENERAL INFORMATION:
;    APPLICANT:  STILLMAN, Bruce
;    APPLICANT:  BELL, Stephen P
;    APPLICANT:  KOBAYASHI, Ryuji
;    APPLICANT:  RINE, Jasper
;    APPLICANT:  FOSS, Margit
;    APPLICANT:  McNALLY, Francis J
;    APPLICANT:  LAURENSON, Patricia
;    APPLICANT:  HERSKOWITZ, Ira
;    APPLICANT:  LI, Joachim J
;    APPLICANT:  GAVIN, Kimberly
;    TITLE OF INVENTION:  ORIGIN OF REPLICATION COMPLEX GENES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,106
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59032/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1676 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  13..1302
US-08-484-106-21


alignment_scores:
           Quality:  105.00              Length:    278
             Ratio:  0.802                 Gaps:     13
 Percent Similarity:  47.122   Percent Identity: 21.942

alignment_block:
 US-09-578-063-47 x US-08-484-106-21   ..

 Align seg 1/1  to: US-08-484-106-21  from: 1  to: 1676

     85 ProLysLysThrGlySerArgProValValLeuGlnHisGlyLeuVa 101
        |||::::|||   |||||||||                  :::     ::
     85 CCAGAAAAAGAAGGATCCAGA.................CAGAAAAAGAC 116

    101 lGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsn..... 114
        ::::|||   :::|||         ||||||   :::|||
    117 GAATGGAAAAGAGAATGCTTCTAGAAATTTGCAATCAAATTTAGAAGAAG 166

    115 ........SerLeuGlyPhe..............IleLeuAlaAspAla 123
                ::::|||||||||                  :::|||::::::
    167 ATTTGGAACAACTGGGCTTCGAGGATGAAACTGTATCAATGGCTCAATCA 216

    124 GlyPheAspValTrp......MetGlyAsnSer............ArgGl 134
        :::::::::   :::      :::|||:::|||              ::
    217 GCAATCGAAAATTACTTTATGCAAGGAAAATCGGCGTCAGAACGAATGAA 266
```

```
    134 yAsnAlaTrpSerArgLysHisLys.................ThrLeuS 145
        :||||||   ||||||:::   :::                  |||   :
    267 TAATGCGAAATCCCGTCGTGGAAGACGTGCTGGAAATGGAAATACTGAAG 316

    145 erIleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMetAlaArg 161
        ::|||:::::::|||:::    :::   :::     ::::::::::::
    317 AAATTGAGGAAGACGATGAGATCAGTAATGCTATCACTGATTTCACAAAA 366

    162 PheAspLeuProAlaValIleAsnPheIleLeuGlnLysThrGlyGlnGl 178
           |||||||||:::::::   |||:::|||   :::|||   :::   ||
    367 TGTGATCTCCCTGGACTTCGAAATTATATTACCAAAAAAGATAACACGGA 416

    178 uLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMetGlyPheIleA 195
        |          :::   :::         :::   :::|||
    417 ATTCGAAAAACGATTGGAGCATCTCGCGGATAATGATTTCGGA....... 459

    195 laPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyrPheAla 211
                                    |||   |||:::|||:::|||
    460 ..........................AAATGGAAGCTTTACCTAGCA 480

    212 LeuAlaProIleAlaThrValLysHisAlaLysSerProGlyThrLysPh 228
           :::
    481 GCTGGA..................................... 486

    228 eLeuLeuLeuProAspMetMetIleLysGlyLeuPheGlyLysLysGluP 245
                     ::::::::::::   |||:::   .:::|||::::::
    487 .........TTTAATATTCTTTTGCACGGTGTCGGTTCGAAGCGTGATG 526

    245 heLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTyrLeuCysGly 261
          |||       |||:::|||   :::::::|||       |||
    527 TTCTC......ACAGAATTTGAGAATGAGCTATCCGATTATACATATATG 570

    262 GlnValIleLeuAspGlnIleCysSerAsnIleMetLeuLeuLeuGlyGl 278
        :::|||          :::          |||:::   :::|||||||||::
    571 AGAGTGGATGCACGGAAAGATGGGCTCAATGTAAAAGTTCTTCTTGGAGC 620

    278 yPheAsnThrAsn.........AsnMetAsnMetSerArgAlaSerValT 292
        ::::|||   |||         |||   |||::::::|||:::
    621 TATCAATGAGAATATGAAGCTGAATTGTAATGTGAAGAGAGGC....... 663

    292 yrAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrp 308
                                       |||:::   :::   |||
    664 ................................CAATCTACGATTAGTTGG 681

    309 SerGlnAlaVal............AsnSerGlyGluLeu........... 317
        ::::::::::::            ||||||   :::|||
    682 GCTCGATCTATTCGCAGAAAAATGAATAGCCAACAGTTGATTCTTATCAT 731

    318 ..........ArgAlaPheAspTrpGlySerGlu 325
                :::|||   ||||||   |||:::
    732 TGATAATATTGAAGCTCCTGATTGGAGAAGTGAT 765

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-602-359A-24

seq_documentation_block:
; Sequence 24, Application US/08602359A
; Patent No. 5942430
;  GENERAL INFORMATION:
;    APPLICANT:  ROBERTSON, Daniel E.
;    APPLICANT:  MURPHY, Dennis
;    APPLICANT:  REID, John
;    APPLICANT:  MAFFIA, Anthony
;    APPLICANT:  LINK, Steven
;    APPLICANT:  SWANSON, Ronald V.
;    APPLICANT:  WARREN, Patrick V.
;    APPLICANT:  KOSMOTKA, Anna
;    TITLE OF INVENTION:  ESTERASES
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & RICHARDSON P.C.
;      STREET:  4225 EXECUTIVE SQUARE, STE 1400
;      CITY:  LA JOLLA
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/602,359A
;      FILING DATE:  February 16, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HAILE, LISA A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09010/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-678-5070
;      TELEFAX:  619-678-5099
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  GENOMIC DNA
US-08-602-359A-24


alignment_scores:
            Quality:  102.00              Length:    180
              Ratio:  1.052                 Gaps:      6
 Percent Similarity:  53.889   Percent Identity: 22.778

alignment_block:
 US-09-578-063-47 x US-08-602-359A-24   ..

 Align seg 1/1  to: US-08-602-359A-24  from: 1  to: 1041

     20 LeuIleLeuValAlaTyrMetPheGlnArgAsnValAsnSerValHisMe 36
        |||:::|||:::|||                  |||   ||||||   ::
    178 CTGCTCCTGGCCGCTGTGGCTGTCTTGGCGCTTGTTGTGTCCGTTGTTGT 227

     36 tProThr............LysAlaValAspProGluAlaPheMetAsnI 49
        :|||:::            :::::::::   |||||||:::        :
    228 CCCGGCTTATGTGGCGTATAGTAGTCTGCACCCGGAGAGCTGTCGGCCCG 277

     49 leSerGluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluVal 65
        :::::               :::|||         :::|||::::::|||
    278 TTGCGCCG............GAGGGGCTCACCTACAAAGAGTTCAGCGTG 315

     66 AlaThrGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLe 82
        ::::::|||||||||   ::::::                  ||||||
    316 ACCGCGGAGGATGGCTTGGTGGTT.................CGGGGCTG 347

     82 uValGlnProLysLysThrGlySerArgProValValLeuLeuGlnHisG 99
         |||            :::|||:::::::||||||   :::|||:::||||
    348 GGTGCTGGGCCCCGGCGCTGGGGGCAACCCGGTGTTCGTTTTGATGCACG 397

     99 lyLeuValGly...GlyAlaSerAsnTrpIleSerAsnLeuProAsnAsn 114
        ||   :::|||      ::::::   :::::::::   |||   ::::::
    398 GGTATACTGGGTGCCGCTCGGCGCCCTACATGGCTGTGCTGGCCCGGGAG 447

    115 SerLeuGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAs 131
                       |||::::::   |||:::   |||   :::   ::
    448 ..............CTCGTGGAGTGGGGGTACCCGGTGGTTGTGTTCGA 482

    131 nSerArgGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspG 148
        :   ||||||::::::   |||            |||::::::
    483 CTTCCGGGGCCACGGGGAGAGCGGGGGCTCGACGACGATTGGG....... 525
```

```
     148 lnAspGluPheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeu 164
                                                 :::|||
     526 ..........................CCCCGGGAGGTGCTGGATGCC 546

     165 ProAlaValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTy 181
         ||||||::::::::::::   ::::::         :::::::|||
     547 CGGGCTGTGGTGGGCTATGTCTCGGAGCGGTTCCCCGGCCGCCGGATAAT 596

     181 rTyrValGlyTyrSerGlnGlyThrThrMetGlyPheIle 194
         ||||||:::|||:::|||   ::::::::::::::::
     597 ATTGGTGGGGTTCAGTATGGGCGGCGCTGTAGCGATCGTG 636

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-232-519-1

seq_documentation_block:
; Sequence 1, Application US/08232519
; Patent No. 5484725
;  GENERAL INFORMATION:
;    APPLICANT:  Kageyama, Bunji
;    APPLICANT:  Nakae, Masanori
;    APPLICANT:  Yagi, Shigeo
;    TITLE OF INVENTION:  NORBORANE TYPE ESTER HYDROLASE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Neave
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,519
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 5-96286
;      FILING DATE:  22-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley Jr, James F
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  SHGN-4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 596-9000
;      TELEFAX:  (212) 596-9090
;      TELEX:  14-8367
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1167 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Acetobacter pasteurianus
;      STRAIN:  ATCC 12873
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1164
US-08-232-519-1


alignment_scores:
            Quality:   98.50            Length:    400
              Ratio:  0.518              Gaps:     20
 Percent Similarity:  47.500   Percent Identity: 20.250

alignment_block:
 US-09-578-063-47 x US-08-232-519-1    ..
```

```
Align seg 1/1  to: US-08-232-519-1  from: 1  to: 1167

      38 ThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGl 54
         |||      ||||||   |||||||||:::
     250 ACGCCAGATGTGGATATTGAAGCCTTTATC..................... 279

      54 nHisGlnGlyTyrProCysGluGluTyrGlu......ValAlaThrGluA 69
          :::::::::   :::|||:::   :::      :::||||||||||
     280 ....AAAAACTTTCATGCAGAAAACCCACAGGATGCGGCTGCAACAGAAG 325

      69 spGlyTyrIleLeuSerValAsnArgIleProArgGly......LeuVal 83
         ||:::         :::   :::::::::      |||         :::
     326 ATGCCTCTGCCGGGGAACCCAAGCAGGTTACGGTAGGCGAACACACGCTA 375

      84 GlnProLysLysThrGlySerArg.........ProValValLeuLeuGl 97
         :::   :::   :::|||::::::         |||:::||||||:::
     376 AATGTGCGTGATGTTGGCACGCAGGAGGGCACGCCCATTGTGCTGGTG.. 423

      97 nHisGlyLeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnA 114
          ||||||:::   |||      |||||||||:::      |||   ::::
     424 .CACGGTTTTGGCGGAGATATCAGCAACTGGCTG......CTCACACAGG 466

     114 snSerLeuGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGly 130
         :::::|||:::                  :::|||||||:::
     467 ATGCCTTGGCCGCAGAAAGGCGCGTAATTGCGTTTGATCTG......... 507

     131 AsnSerArgGlyAsnAlaTrpSerArgLysHis.........LysThrLe 144
                |||::::::   |||   |||:::           |||||
     508 ......CCGGGGCATGGGGCTTCCTCTAAAAACGTGGGCACAGGCACGCT 551

     144 uSerIleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMetAlaA 161
         |::::::
     552 GGCGTTT.......................................... 558

     161 rgPheAspLeuProAlaValIleAsnPheIleLeuGlnLysThrGlyGln 177
                |||   :::|||::::::   :::||||||
     559 ........TTGGCCGGTGTGGTAAGCGAATTGCTGCAAACCCTTAAAATA 600

     178 GluLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMetGlyPheIl 194
         ||||||   :::   ||||||:::|||   |||          |||   ||
     601 GAAAAAGCCCATGTGGTGGGCCATTCTTTGGGG.........GGCGGCAT 641

     194 eAlaPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyrPheA 211
         ||||::::::   :::   :::       :::::::::
     642 TGCCCTGACCCTGCTGCGAGATCACCCTGATCAGGTTGCCAGCCTGAACC 691

     211 laLeuAlaProIleAlaThrValLysHisAlaLysSerProGlyThrLys 227
           |||||||||   :::             :::|||:::   ::::::
     692 TTTTGGCCCCAGCCGGGTTG.........GGTAAGGATGTGAATGCAGAT 732

     228 PheLeuLeuLeuProAspMetMetIleLysGlyLeuPheGlyLysLysGl 244
         |||:::                        :::   |||      :::||
     733 TTTATC.....................AGCGCATTTGTGGATAGTGA 758

     244 uPheLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTyr....... 258
         |          :::   :::   :::   |||:::::::::|||
     759 AAGCAGCCGCGATATGAAGGCTGTTTTGCAAATGCTGGTGTATAACAAAG 808

     259 ..LeuCysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMetLeu 274
           |||   |||:::   ::::::|||   :::
     809 CCCTAGTGGGCCGTAAGATGGTGGATGCCGTG.................. 840

     275 LeuLeuGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaSerVa 291
                 ::::::::::::::::::::   :::|||
     841 ............CTGCGTGCACGTAGGCTAGATGGCGCGCGGGAT..... 873

     291 lTyrAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisT 308
             |||   |||:::::::|||   :::          |||
     874 :...GCCCTGCACGTTATTGCTAAAGCGTGCTTCCCCAAC.......... 909

     308 rpSerGlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGlySer 324
                 :::   ::::::   :::|||::::::         |||:::
```

```
 910 ........GGGCATCAGGCGGATGATCTGCACTCGGTGCTAGCTGGGGCG 951

 325 GluThrLysAsnLeuGluLysCysAsnGlnProThrProValArgTyrAr 341
     ||||||
 952 GAAACA..................................................... 957

 341 gValArgAspMetThrValProThrAlaMetTrpThrGlyGlyGlnAspT 358
                      ||||||   ::::::   |||   :::|||
 958 ..................CCTACCCAGATTTTCTGGGGCAAGGAAGATG 988

 358 rpLeuSerAsnProGluAspValLysMetLeuLeuSerGluValThrAsn 374
       :::   :::   :::::::::      |||   :::   :::
 989 AAATTCTTTCTGTCTCCAACGCCGCTGGCCTGCCAGATGTCATCCCCGTG 1038

 375 LeuIleTyrHisLys......AsnIleProGluTrpAlaHisValAspPh 389
        :::|||::::::      ::::::|||:::      |||:::
1039 ACAGTGTATGAAGAAACAGGCCATCTGCCGCAGCTTGAACATGCAACAGA 1088

 389 eIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGluIleIleHisL 406
      :::                        .      |||:::   |||   |
1089 TGTG..........................AACAAAGCCATTGCCC 1108

 406 eu...MetGlnGlnGluGluThrAsnLeuSerGlnGlyArgCysGluAla 421
     ||   :::::::::   |||:::   ||||||::::::|||   :::|||
1109 TGTTTGTAAAAGACCCCGAAGCCGCGCTGAGCATGGCCCGGATGGACGCG 1158
```

```
OM of: US-09-578-063-47 to: EST:*   out_format : pfs

Date: Jan 30, 2002 11:35 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09578063/runat_30012002_092150_26299/app_query.fasta_1.488
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09578063_@CGN1_1_3861 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-578-063-47
Query length: 423
Database: EST:*
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 1532.800000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_htc:AK019504       + 1836.00 3644.64  7.6e-194  2927  ! AK019504 Mus musculus 0 day ne
gb_htc:AK009571       + 1139.50 2256.79  1.5e-116  1349  ! AK009571 Mus musculus adult ma
gb_htc:AK009875       + 1139.50 2256.79  1.5e-116  1349  ! AK009875 Mus musculus adult ma
gb_htc:AK010093       + 1139.50 2256.78  1.5e-116  1350  ! AK010093 Mus musculus adult ma
gb_htc:AK010139       + 1139.50 2256.78  1.5e-116  1350  ! AK010139 Mus musculus adult ma
gb_htc:AK009431       + 1139.50 2256.76  1.5e-116  1352  ! AK009431 Mus musculus adult ma
gb_htc:AK009479       + 1138.50 2254.83  2.0e-116  1343  ! AK009479 Mus musculus adult ma
gb_htc:AK009473       + 1138.50 2254.81  2.0e-116  1345  ! AK009473 Mus musculus adult ma
gb_htc:AK009523       + 1138.50 2254.80  2.0e-116  1346  ! AK009523 Mus musculus adult ma
gb_htc:AK009729       + 1138.50 2254.80  2.0e-116  1346  ! AK009729 Mus musculus adult ma
gb_htc:AK010116       + 1138.50 2254.80  2.0e-116  1347  ! AK010116 Mus musculus adult ma
gb_htc:AK010058       + 1138.50 2254.79  2.0e-116  1348  ! AK010058 Mus musculus adult ma
gb_htc:AK009300       + 1138.50 2254.78  2.0e-116  1349  ! AK009300 Mus musculus adult ma
gb_htc:AK009474       + 1138.50 2254.78  2.0e-116  1349  ! AK009474 Mus musculus adult ma
gb_htc:AK009546       + 1138.50 2254.78  2.0e-116  1349  ! AK009546 Mus musculus adult ma
gb_htc:AK009573       + 1138.50 2254.78  2.0e-116  1349  ! AK009573 Mus musculus adult ma
gb_htc:AK009459       + 1138.50 2254.77  2.0e-116  1350  ! AK009459 Mus musculus adult ma
gb_htc:AK009525       + 1138.50 2254.77  2.0e-116  1350  ! AK009525 Mus musculus adult ma
gb_htc:AK009773       + 1138.50 2254.77  2.0e-116  1350  ! AK009773 Mus musculus adult ma
gb_htc:AK010061       + 1138.50 2254.77  2.0e-116  1350  ! AK010061 Mus musculus adult ma
gb_htc:AK010125       + 1138.50 2254.77  2.0e-116  1350  ! AK010125 Mus musculus adult ma
gb_htc:AK009428       + 1138.50 2254.76  2.0e-116  1351  ! AK009428 Mus musculus adult ma
gb_htc:AK010019       + 1138.50 2254.72  2.0e-116  1356  ! AK010019 Mus musculus adult ma
gb_htc:AK010035       + 1138.50 2254.72  2.0e-116  1356  ! AK010035 Mus musculus adult ma
gb_htc:AK010203       + 1136.50 2250.76  3.3e-116  1350  ! AK010203 Mus musculus adult ma
gb_htc:AK010124       + 1135.50 2248.84  4.3e-116  1340  ! AK010124 Mus musculus adult ma
gb_htc:AK009413       + 1135.50 2248.80  4.3e-116  1345  ! AK009413 Mus musculus adult ma
gb_htc:AK010236       + 1135.50 2248.78  4.3e-116  1347  ! AK010236 Mus musculus adult ma
gb_htc:AK009990       + 1134.50 2246.76  5.6e-116  1349  ! AK009990 Mus musculus adult ma
gb_htc:AK009537       + 1133.50 2244.75  7.2e-116  1350  ! AK009537 Mus musculus adult ma
gb_htc:AK009359       + 1132.50 2242.78  9.3e-116  1346  ! AK009359 Mus musculus adult ma
gb_htc:AK009544       + 1131.50 2240.78  1.2e-115  1345  ! AK009544 Mus musculus adult ma
gb_htc:AK010103       + 1130.50 2238.76  1.6e-115  1347  ! AK010103 Mus musculus adult ma
gb_htc:AK009560       + 1129.50 2236.76  2.0e-115  1346  ! AK009560 Mus musculus adult ma
gb_htc:AK010026       + 1129.50 2236.74  2.0e-115  1349  ! AK010026 Mus musculus adult ma
gb_htc:AK010231       + 1128.50 2234.74  2.6e-115  1348  ! AK010231 Mus musculus adult ma
gb_htc:AK009437       + 1127.50 2232.75  3.3e-115  1346  ! AK009437 Mus musculus adult ma
gb_htc:AK010036       + 1125.50 2228.73  5.6e-115  1348  ! AK010036 Mus musculus adult ma
gb_htc:AK010148       + 1124.50 2226.72  7.3e-115  1349  ! AK010148 Mus musculus adult ma
gb_htc:AK010106       + 1121.50 2220.70  1.6e-114  1349  ! AK010106 Mus musculus adult ma
gb_htc:AK010173       + 1120.50 2218.72  2.0e-114  1346  ! AK010173 Mus musculus adult ma
gb_htc:AK010123       + 1110.50 2198.64  2.7e-113  1350  ! AK010123 Mus musculus adult
gb_htc:AK009923       + 1109.50 2196.19  3.6e-113  1407  ! AK009923 Mus musculus adult
gb_htc:AK010091       + 1086.50 2150.56  1.3e-110  1347  ! AK010091 Mus musculus adult
gb_htc:AK010110       + 1075.50 2128.46  2.2e-109  1353  ! AK010110 Mus musculus adult

seq_name: gb_htc:AK019504

seq_documentation_block:
LOCUS       AK019504     2927 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus 0 day neonate skin cDNA, RIKEN full-length enriched
            library, clone:4632427C23, full insert sequence.
ACCESSION   AK019504
VERSION     AK019504.1  GI:12859754
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 0 day neonate skin cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:4632427C23.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2927)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 2927)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 2927)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 2927)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2927)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
```

further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 3 and 7 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

```
FEATURES             Location/Qualifiers
     source          1. .2927
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1912763"
                     /db_xref="MGD:MGI:1926003"
                     /clone="4632427C23"
                     /tissue_type="skin"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="0 day neonate"
     CDS             153. .662
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB31766.1"
                     /db_xref="GI:12859755"
                     /translation="MSEILSRVWTVSHRVEIWLLILVAYLLQRNVNSGHLPTKAADPE
                     AFMNVSEIIKHKGYPSEEYEVATEDGYILSVNRIPRGQTRLKKEGSRPVVLLQHGLLG
                     DASNWISNLPNNSLGFILADAGFDVWMGNSRGNTWSRKHKTLSIDQDEFWAFRYKNTL
                     RVEVDTDMP"
     polyA_signal    2908. .2913
                     /note="putative"
     polyA_site      2927
                     /note="putative"
BASE COUNT      837 a    581 c    664 g    845 t
ORIGIN
```

```
alignment_scores:
            Quality: 1836.00              Length:    682
              Ratio:  4.567                Gaps:        2
 Percent Similarity:  58.944   Percent Identity: 54.106

alignment_block:
 US-09-578-063-47 x AK019504   ..

 Align seg 1/1  to: AK019504  from: 1  to: 2927

       1 MetLeuGluThrLeuSerArgGlnTrpIleValSerHisArgMetGluMe 17
         |||   |||   |||||||||   |||   ||||||||||||:::|||::
     153 ATGTCAGAAATCTTGTCAAGAGTGTGGACTGTTTCGCACAGAGTGGAGAT 202

      17 tTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAsnServ 34
         :|||||||||||||||||||||||:::::|||||||||||||||||||
     203 ATGGCTTCTGATTCTGGTAGCGTATTTACTCCAAAGAAATGTGAACTCGG 252

      34 alHisMetProThrLysAlaValAspProGluAlaPheMetAsnIleSer 50
           |||:::|||||||||||||:::|||||||||||||||||||||:::|||
     253 GACATTTGCCCACGAAAGCTGCGGATCCAGAAGCATTCATGAATGTTAGC 302

      51 GluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaTh 67
         |||||||||:::|||:::|||||||||   ||||||||||||||||||
     303 GAAATCATCAAACACAAGGGTTATCCCAGTGAGGAGTATGAAGTTGCAAC 352

      67 rGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuValG 84
         ||||||||||||||||||||||||||||||||||||||||   ::::
```

```
    353 GGAAGATGGGTACATCCTTTCTGTGAACAGAATCCCTCGGGGACAGACAC 402

     84 lnProLysLysThrGlySerArgProValValLeuLeuGlnHisGlyLeu 100
        ::   ||||||   ||||||||||||||||||||||||||||||||||||
    403 GGTTAAAGAAGGAAGGATCCAGGCCAGTGGTGTTACTGCAACATGGTCTT 452

    101 ValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGl 117
        :::|||   ||||||||||||||||||||||||||||||||||||||||
    453 CTTGGGGATGCTAGCAACTGGATTTCCAACCTGCCCAACAACAGCCTGGG 502

    117 yPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgG 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    503 TTTTATTCTGGCAGATGCAGGTTTTGATGTGTGGATGGGAAACAGCAGAG 552

    134 lyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAspGlu 150
        |||||:::||||||||||||||||||||||||||||||||||||||||||
    553 GAAACACCTGGTCTCGGAAGCACAAGACCCTCTCCATAGATCAAGATGAG 602

    151 PheTrpAlaPhe..................................... 154
        ||||||||||||
    603 TTCTGGGCTTTCAGGTACAAGAACACTTTGAGGGTGGAGGTAGATACCGA 652

    154 ..... 154

    653 CATGCCCTGAAGCACTGGGTAAAGAATTGTGACTTGTGGTATTTGGATTG 702

    154 ..... 154

    703 CTGTGTAGTTCCCAACTCAATAACCCCCAAGTTTTCTTATTATCCTGTGC 752

    154 ..... 154

    753 TGATGTGGGTGAGTGTGGGGAAGACTGTGTCTCCCAAGGTCTAGCTAGTG 802

    154 ..... 154

    803 GGTTGAAAACAGTGTCTGGTGTCAGGTGATTGAGAATGGAGTGGGAAAGG 852

    154 ..... 154

    853 TATCCAGTCATTTCCGTGTGAGTGAGTATGTAGGATTAACTCCCTCGTAT 902

    154 ..... 154

    903 AGGACAACAGGACTGCCACCCCCAAGACATGAGACCAAGTGTCATTAAAT 952

    154 ..... 154

    953 AGGGATTTAACAATCACTGTGTGTCGTCACCTTAAATAATATTTCCCATA 1002

    154 ..... 154

   1003 GCTTTACTCAAGTTCCTAATGTCCTTACTAACCACTGATGGTCAACCTGT 1052

    154 ..... 154

   1053 TCTCTATGCATTTGGTCAGATATGAGTTTTGTTACCATTCATCCTGCAAA 1102

    154 ..... 154

   1103 GAAATCTGCATACCCCTGAGCTGTTTTAGTTGCAGGGTAGGACATGCTTC 1152

    154 ..... 154

   1153 ACATCAGCAGTTGGCTGGTGAGTTTAAAACTTCTCAATAACCATTACTAT 1202

    154 ..... 154

   1203 AGTACAGCATCCAGAGAGAATAGACATTTCTCAGGTTTGAAACAAATCCG 1252

    154 ..... 154

   1253 TCTCAAGGATTTTGAGGAACACGGTGACAAACTAACAACACTGACACTGG 1302
```

```
 154 ..................................................  154

1303 TCGGGAGGACTTGTTAGCAGATCCTGGGGGATTCTGACCATGTTCTGTTT 1352

 155 .....................................SerTyrAspGl  158
                                                 ||||||||||
1353 GTTTTATCATTGTCGAATAACTTTGTCTTTTCCCTTTGTAGTTATGATGA 1402

 158 uMetAlaArgPheAspLeuProAlaValIleAsnPheIleLeuGlnLysT  175
      ||||||||||||||||||||||||||||||||||||||||||||||||||
1403 AATGGCTAGGTTTGACCTTCCAGCTGTGATAAACTTTATCTTACAGAAAA 1452

 175 hrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMet  191
      ||||||||:::|||:::|||||||||||||||||||||||||||||||||
1453 CGGGCCAGAAAAAGGTCTATTATGTCGGCTACTCACAGGGCACCACCATG 1502

 192 GlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLysIleLysMe  208
      ||||||||||||||||||||||||||||||||:::|||||||||||||
1503 GGCTTTATTGCATTTTCCACAATGCCAGAGCTAGCTCATAAAATCAAAAT 1552

 208 tTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLysSerProG  225
      ||||||||||||||||||||||||||||||||:::|||:::|||||||||
1553 GTATTTTGCCTTAGCTCCTATAGCCACTGTTAAATATGCAAGAAGTCCTG 1602

 225 lyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyLeuPheGly  241
      ||||||||||||||||||||||||||||||||||||   |||||||||
1603 GTACCAAATTTCTGCTGCTGCCAGATATGATGATCAAGGTATTATTTGGC 1652

 242 LysLysGluPheLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTy  258
      ::::::|||||||||||||||||||||||||:::|||||||||   |||||
1653 AGACAAGAGTTTTTATACCAGACTAGATTTTTCAGACAGCTTTTTATTTA 1702

 258 rLeuCysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMetLeuL  275
      ||||||||||||:::|||||||||||||||||||||||||||:::|||||
1703 CCTTTGTGGCCAGATGATTCTTGACCAAATCTGCAGCAACATCATCTTAC 1752

 275 euLeuGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaSerVal  291
      ||||||||||||||||||||||||||||||||||||||||||||:::|||
1753 TCCTGGGAGGTTTTAACACAAACAATATGAACATGAGCAGAGCAAATGTG 1802

 292 TyrAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTr  308
      |||:::||||||||||   ||||||||||||||||||||||||||||||
1803 TATGTCGCCCATACACCTGCTGGAACGTCTGTGCAGAACATTCTCCACTG 1852

 308 pSerGlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGlySerG  325
      ||||||||||||||||||||||||||||||||||||||||||||||||||
1853 GAGCCAGGCCGTGAATTCTGGGGAACTTCGTGCCTTTGACTGGGGAAGTG 1902

 325 luThrLysAsnLeuGluLysCysAsnGlnProThrProValArgTyrArg  341
      ||||||||||   ||||||||||||||||||||||:::|||||||:::
1903 AGACCAAAAATCAGGAGAAATGTAATCAGCCAACTCCTATAAGGTACAAA 1952

 342 ValArgAspMetThrValProThrAlaMetTrpThrGlyGlyGlnAspTr  358
      ||||||||||||   ||||||||||||||||||||||||||||||||||
1953 GTTCGAGATATGATGGTCCCAACAGCAATGTGGACTGGAGGTCAAGACTG 2002

 358 pLeuSerAsnProGluAspValLysMetLeuLeuSerGluValThrAsnL  375
      ||||||||||||:::|||||||||   ||||||||||||||||||||||
2003 GCTTTCAAATCCAGATGATGTGAAAACATTACTTTCTGAAGTAACCAACC 2052

 375 euIleTyrHisLysAsnIleProGluTrpAlaHisValAspPheIleTrp  391
      ||||||||||||||||||||||||||||||||||||||||||||||||||
2053 TCATCTACCACAAGAACATTCCTGAGTGGGCTCATGTGGATTTCATCTGG 2102

 392 GlyLeuAspAlaProHisArgMetTyrAsnGluIleIleHisLeuMetGl  408
      ||||||||||||||||:::|||:::|||||||||||||||||||||||::
2103 GGGCTGGATGCCCCTCAGCGTGTTTACAATGAAATTATACATCTGATGAA 2152

 408 nGlnGluGluThrAsnLeuSerGlnGlyArgCysGluAlaValLeu 423
      :|||   |||   ||||||   ||||||   |||::::::   |||
2153 GCAG...GAGCCCAACCTTCCCCAGGGAACCTGCAGGGTCAAATTG 2195
```

```
seq_name: gb_htc:AK009571

seq_documentation_block:
LOCUS       AK009571     1349 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310032G08, full insert sequence.
ACCESSION   AK009571
VERSION     AK009571.1  GI:12844447
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310032G08.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1349)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1349)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1349)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1349)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1349)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
```

prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

```
FEATURES             Location/Qualifiers
     source          1. .1349
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1895763"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310032G08"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             52. .1239
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26368.1"
                     /db_xref="GI:12844448"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      396 a    290 c    284 g    379 t
ORIGIN
```

```
alignment_scores:
           Quality: 1139.50              Length:    411
             Ratio:  3.485                Gaps:        4
 Percent Similarity:  79.562   Percent Identity: 52.555

alignment_block:
 US-09-578-063-47 x AK009571   ..

 Align seg 1/1  to: AK009571  from: 1  to: 1349

       7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
         |||:::|||           :::::::::||||||||||||:::::::
      34 AGGCACTGG...........CACTTCAAGATGTGGCTGCTATTAGTAAC 71

      23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
         ::::   :::           ::::::::::::::
      72 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 121

      39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
              :::|||||||||   ||||||:::|||:::::|||   :::
     122 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 171

      56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
         |||||||||   ||||||||||||||:::||||||||||||||||||
     172 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 221

      72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
         ||||:::|||   |||||||||   |||      :::   ::::::   |
     222 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 271

      89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
         ||:::||||||||:::   |||||||||||||||:::::::::|||:::
     272 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 321

     106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
         |||||||||:::|||||||||||||||||||:::|||||||||||||||
     322 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 371

     122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
         |||||||:::|||||||||:::|||||||||||||||||||:::|||||||
     372 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 421

     139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
         |||||:::          |||   |||:::   ||||||||||||||||||
     422 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 471

     156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
         :::|||||||||||||:::::|||||||||||:::|||:::||||||::
     472 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 521

     172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
         :||||||||||||||||||||||||:::||||||||||:::||||||||||
     522 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 571

     189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
         |||||:::|||||||||||||||||||||   |||   ||||||:::|||
     572 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 621

     206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
         ||||||  ::::::|||||||||||||:::|||||||||||||:::::::
     622 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 671

     222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
         :||||||      |||:::.  |||:::|||   ::::::::|||   :
     672 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 721

     239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
         ::||||||:::|||   |||:::         |||      :::||||||
     722 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 771

     253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
                |||         |||::::::   :::|||||   :::||
     772 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 812

     269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
         |||||||   ::::::::::::   |||||||:::   :::|||:::|||:
     813 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 862

     286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
         ::||||||   :::||||||   :::|||:::   |||||||||||:::
     863 TGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 912

     303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
         |||:::::::::|||||:::|||   ::::::||||||:::|||:::||
     913 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 962

     319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
         |:::::||||||||||      :::|||:::         ||||||   |
     963 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1012

     336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
         |||||      |||   |||      ||||||||||||   |||:::|||
    1013 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1062

     353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
         :::||||||:::|||   |||:::::::|||:::||||||   ||||||||
    1063 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1112

     369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
         |   ::::::   ||||||:::||||||||:::|||      :::   |
    1113 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1162

     386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
         ||:::||||||||||||:::::|||||||||:::::::::|||||||||
    1163 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1212

     403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
```

```
        |||:::   :::|||  ::::::      |||
    1213 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1245

seq_name: gb_htc:AK009875

seq_documentation_block:
LOCUS       AK009875     1349 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310047G15, full insert sequence..
ACCESSION   AK009875
VERSION     AK009875.1  GI:12844938
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310047G15.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1349)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1349)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1349)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1349)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1349)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1349
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1898366"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310047G15"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              53. .1240
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26556.1"
                      /db_xref="GI:12844939"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLRNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      396 a    290 c    284 g    379 t
ORIGIN


alignment_scores:
           Quality: 1139.50            Length:    411
             Ratio:  3.495               Gaps:      4
 Percent Similarity:  79.319   Percent Identity: 52.555

alignment_block:
 US-09-578-063-47 x AK009875   ..

 Align seg 1/1  to: AK009875  from: 1  to: 1349

       7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuIleLeuVa 23
         |||:::|||            :::   :::|||||||||||||:::::::
      35 AGGCACTGG...........CACTACAAGATGTGGCTGCTATTAGTAAC 72

      23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
         ::::  :::           ::::::::::::::
      73 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 122

      39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
            :::|||||||||   ||||||:::|||:::::|||   :::
     123 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 172

      56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
         |||||||||   ||||||||||||||:::||||||||||||||||||||
     173 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 222

      72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
         ||||:::|||   |||||||||   |||      :::   ::::::   |
     223 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 272

      89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
```

```
273 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 322

106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
323 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 372

122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
373 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 422

139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
423 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 472

156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
473 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 522

172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
523 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 572

189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
573 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 622

206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
623 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 672

222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
673 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 722

239 euPheGlyLysLysGluPheLeu.....TyrGlnThrArgPheLeu 252
723 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 772

253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
773 GGTACGGAAGTG.....TGCTCACGGGAGCTGCTAGATCTTCTCTG 813

269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
814 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 863

286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
864 TGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 913

303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
914 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 963

319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
964 CTATAACTGGGGAAGTCCATTACGGAACATGTTACACTACAATCAGAAAA 1013

336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
1014 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1063

353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
1064 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1113

369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
1114 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1163

386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
```

```
1164 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1213

 403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
1214 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1246
```

seq_name: gb_htc:AK010093

seq_documentation_block:

LOCUS AK010093 1350 bp mRNA HTC 05-JUL-2001

DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310067K20, full insert sequence.

ACCESSION AK010093

VERSION AK010093.1 GI:12845297

KEYWORDS CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2310067K20.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1350)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE 2 (bases 1 to 1350)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE 3 (bases 1 to 1350)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome research. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE 4 (bases 1 to 1350)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5 (bases 1 to 1350)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

```
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1350
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1898392"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310067K20"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53. .1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26697.1"
                     /db_xref="GI:12845298"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLVASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      396 a    290 c    286 g    378 t
ORIGIN


alignment_scores:
            Quality: 1139.50              Length:    411
              Ratio:  3.495                Gaps:        4
 Percent Similarity:  79.319   Percent Identity: 52.798

alignment_block:
 US-09-578-063-47 x AK010093   ..

 Align seg 1/1  to: AK010093  from: 1  to: 1350

       7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
         |||:::|||           :::   :::|||||||||||||:::::::
      35 AGGCACTGG...........CACTACAAGATGTGGCTGCTATTAGTAAC 72

      23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
         ::::   :::           ::::::::::::::
      73 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 122

      39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
            :::||||||||||   ||||||:::|||:::::::|||   :::
     123 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 172

      56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
         |||||||||   |||||||||||||||::|||||||||||||||||||
     173 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 222

      72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
         ||||:::|||   |||||||||   |||      :::   ::::::   |

     223 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 272

      89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
         ||:::||||||||:::   |||||||||||||||||||:::::|||:::
     273 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGGTTGCATCAGCCACA 322

     106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
         ||||||||||:::|||||||||||||||||||:::|||||||||||||
     323 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 372

     122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
         |||||||:::|||||||||:::|||||||||||||||||||:::|||||||
     373 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 422

     139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
         |||||:::     .   |||   |||:::   ||||||||||||||||||
     423 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 472

     156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
         :::|||||||||||:::::|||||||||||:::|||:::|||||||::
     473 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 522

     172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
         :|||||||||||||||||||||||:::|||||||||:::||||||||||
     523 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 572

     189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
         |||||:::|||||||||||||||||||||   |||   ||||||:::|||
     573 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 622

     206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
         ||||||   ::::::|||||||||||:::|||||||||||:::::::::
     623 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 672

     222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
         :||||||      |||:::   |||:::|||   ::::::::::|||   :
     673 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 722

     239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
         ::||||||:::|||   |||:::         |||      :::||||||
     723 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 772

     253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
                |||         |||::::::   :::||||||   :::||
     773 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 813

     269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
         |||||||   ::::::::::::   |||||||:::   :::|||:::|||:
     814 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 863

     286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
         ::||||||   :::||||||   :::|||:::   |||||||||||||:::
     864 TGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 913

     303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
         |||::::::::::||||||:::|||   ::::::||||||:::|||:::||
     914 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 963

     319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
         |::::::||||||||      :::|||:::          ||||||   |
     964 CTATAACTGGGGGAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1013

     336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
         |||||      |||   |||      ||||||||||||   |||:::|||
    1014 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1063

     353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
         :::||||||:::|||   |||:::::::|||:::||||||   ||||||||
    1064 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1113

     369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
         |   ::::::   ||||||:::||||||||:::|||      :::   |
    1114 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1163
```

```
     386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
         ||:::|||||||||||||:::::::|||||||||:::::::::|||||||||
    1164 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1213

     403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
         |||:::   :::|||   ::::::      |||
    1214 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1246

seq_name: gb_htc:AK010139

seq_documentation_block:
LOCUS       AK010139     1350 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310069P19, full insert sequence.
ACCESSION   AK010139
VERSION     AK010139.1  GI:12845371
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310069P19.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1350)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1350)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1350)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1350)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1350)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'.
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1350
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1898377"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310069P19"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              53. .1240
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26725.1"
                      /db_xref="GI:12845372"
                      /translation="MWLLLYTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPIATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      398 a    289 c    283 g    380 t
ORIGIN


alignment_scores:
           Quality: 1139.50            Length:    411
             Ratio:  3.495              Gaps:      4
 Percent Similarity:  79.319   Percent Identity: 52.798

alignment_block:
 US-09-578-063-47 x AK010139   ..

 Align seg 1/1  to: AK010139  from: 1  to: 1350

       7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
         |||:::|||            :::   :::||||||||||||:::::::
      35 AGGCACTGG............CACTACAAGATGTGGCTGCTATTAGTAAC 72

      23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
         ::::  :::          ::::::::::::::
      73 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 122

      39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
               :::||||||||||   ||||||:::|||::::::|||   :::
     123 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 172

      56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
          |||||||||   ||||||||||||||:::||||||||||||||||||
     173 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 222
```

```
 72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
    ||||:::|||   ||||||||   |||      :::   ::::::   |
223 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 272

 89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
    ||:::||||||||||:::   ||||||||||||||:::::::::|||:::
273 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 322

106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
    |||||||||:::|||||||||||||||||||:::|||||||||||||||
323 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 372

122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
    |||||||:::|||||||||:::|||||||||||||||||||:::|||||||
373 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 422

139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
    |||||:::         |||   |||:::   ||||||||||||||||||
423 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 472

156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
    :::|||||||||||||:::::|||||||||||||:::|||:::|||||||::
473 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 522

172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
    :|||||||||||||||||||||||:::||||||||::::||||||||||
523 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 572

189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
    |||||:::|||||||||||||||||||||   |||   ||||||:::|||
573 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 622

206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
    ||||||   ::::::|||||||||||||||||||||||||||||:::::::
623 ATCAAGAGGTTTTATGCATTAGCTCCAATTGCTACTGTGAAGTATACAGA 672

222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
    :||||||      |||:::   |||:::|||   ::::::::::|||   :
673 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 722

239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
    ::||||||:::|||   |||:::         |||      :::|||||||
723 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 772

253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
          |||         |||:::::   :::||||||   :::||
773 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 813

269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
    |||||||   ::::::::::::   |||||||:::   :::|||:::|||:
814 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 863

286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrServal 302
    ::|||||||   :::||||||   :::|||:::   |||||||||||||:::
864 TGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 913

303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
    |||:::::::::|||||:::|||   :::::::|||||:::|||:::||
914 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 963

319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
    |:::::::|||||||||      :::|||:::          ||||||   |
964 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1013

336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
    |||||      |||   |||      ||||||||||||   |||:::|||
1014 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1063

353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
    :::|||||:::|||   |||:::::|||:::|||||||   |||||||||
1064 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1113

369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
    |   ::::::   ||||||:::||||||||||:::|||      :::   |
1114 TCCCAAACTCCCCAACCTTTTGTACCATAAGGAGATTCTTCCCTACAATC 1163

386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
    ||:::|||||||||||||:::::::||||||||:::::::::|||||||||
1164 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1213

403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
    |||:::   :::|||   ::::::      |||
1214 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1246
```

seq_name: gb_htc:AK009431

seq_documentation_block:

```
LOCUS       AK009431     1352 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310021B16, full insert sequence.
ACCESSION   AK009431
VERSION     AK009431.1  GI:12844222
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310021B16.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1352)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1352)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1352)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1352)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1352)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
```

```
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1352
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1894047"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310021B16"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              55. .1242
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26283.1"
                      /db_xref="GI:12844223"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRLDV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      397 a    291 c    285 g    379 t
ORIGIN


alignment_scores:
           Quality: 1139.50            Length:    411
             Ratio:  3.495              Gaps:      4
 Percent Similarity:  79.319   Percent Identity: 52.555

alignment_block:
 US-09-578-063-47 x AK009431   ..

 Align seg 1/1  to: AK009431  from: 1  to: 1352

       7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
         |||:::|||          :::   :::|||||||||||||:::::::::
      37 AGGCACTGG...........CACTACAAGATGTGGCTGCTATTAGTAAC 74

      23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
         ::::   :::           ::::::::::::::
      75 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 124

      39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
               :::|||||||||   ||||||:::|||:::::::|||   :::
     125 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 174

      56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
            |||||||||   ||||||||||||||||:::|||||||||||||||||
     175 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 224

      72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
         ||||:::|||   |||||||||   |||       :::   ::::::   |
     225 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 274

      89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
         ||:::||||||||||:::   ||||||||||||||:::::::::|||:::
     275 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 324

     106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
         ||||||||||:::|||||||||||||||||||||:::|||||||||||||
     325 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 374

     122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
         |||||||:::|||||||||:::||||||||||||||||||:::|||||||
     375 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 424

     139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
         |||||:::          |||   |||:::   ||||||||||||||||||
     425 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 474

     156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
         :::|||||||||||||:::::::||||||||||||:::|||:::|||||::
     475 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 524

     172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
         :||||||||||||||||||||||:::|||||||||:::||||||||||
     525 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 574

     189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
         |||||:::|||||||||||||||||||||||  |||   ||||||:::|||
     575 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 624

     206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
         ||||||   ::::::||||||||||||:::||||||||||||:::::::
     625 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 674

     222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
         :||||||   |||:::   |||:::|||   ::::::::::|||   :
     675 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 724

     239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
         ::||||||:::|||   |||:::       |||      :::||||||
     725 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 774

     253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
               |||         |||::::::  :::||||||   :::||
     775 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 815

     269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
         |||||||   ::::::::::::   |||||||:::   :::|||:::|||:
     816 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 865

     286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
         ::||||||   :::||||||   :::|||:::   ||||||||||||:::
     866 TGAGTCGCCTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 915

     303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
         |||:::::::::|||||||:::|||   :::::|||||:::|||:::||
     916 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 965

     319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
         |:::::||||||||||      :::|||:::          ||||||   |
     966 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1015

     336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
         |||||   |||   |||      ||||||||||||   |||:::|||
    1016 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1065

     353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
```

```
      :::|||||||:::|||   |||:::::::|||:::||||||   ||||||||
   1066 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1115

    369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
        |   ::::::   ||||||:::|||||||||:::|||      :::   |
   1116 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1165

    386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
        ||:::|||||||||||||:::::|||||||||:::::::::|||||||||
   1166 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1215

    403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
        |||:::   :::|||   ::::::      |||
   1216 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1248

seq_name: gb_htc:AK009479

seq_documentation_block:
LOCUS       AK009479     1343 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310022K08, full insert sequence.
ACCESSION   AK009479
VERSION     AK009479.1  GI:12844303
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310022K08.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1343)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1343)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1343)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1343)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1343)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1343
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1896992"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310022K08"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             50. .1237
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26316.1"
                     /db_xref="GI:12844304"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEIPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      394 a    289 c    282 g    378 t
ORIGIN


alignment_scores:
            Quality: 1138.50            Length:    411
              Ratio:  3.492              Gaps:        4
 Percent Similarity:  79.319   Percent Identity: 52.555

alignment_block:
 US-09-578-063-47 x AK009479   ..

 Align seg 1/1  to: AK009479  from: 1  to: 1343

       7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
         |||:::|||           :::   :::|||||||||||||:::::::::
      32 AGGCACTGG...........CACTACAAGATGTGGCTGCTATTAGTAAC 69

      23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
         ::::   :::           ::::::::::::::::
      70 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 119

      39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
```

```
          :::|||||||||   ||||||:::|||::::::|||   :::
  120 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 169

   56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
      |||||||||  ||||||||||||||:::||||||||||||||||||||
  170 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 219

   72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
      ||||:::|||  |||||||||  |||     :::   ::::::   |
  220 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 269

   89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
      ||:::||||||||:::  ||||||||||||||||:::::::::|||:::
  270 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 319

  106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
      |||||||||:::|||||||||||||||||||||:::|||||||||||||
  320 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 369

  122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
      |||||||:::|||||||||:::|||||||||||||||||||:::|||||||
  370 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 419

  139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
      |||||:::         |||   |||:::   ||||||||||||||||||
  420 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 469

  156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
      :::|||||||||||||:::::|||||||||||||:::|||:::|||||::
  470 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 519

  172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
      :|||||||||||||||||||||||||:::||||||||||:::||||||||||
  520 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 569

  189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
      |||||:::|||||||||||||||||||||   |||   |||||:::|||
  570 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 619

  206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
      ||||||   ::::::|||||||||||||:::|||||||||||||:::::::
  620 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 669

  222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
      :||||||     |||:::   |||:::|||   ::::::::::|||   :
  670 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 719

  239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
      ::|||||||:::|||   |||:::         |||      :::||||||
  720 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 769

  253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
          |||         |||::::::   :::||||||   :::||
  770 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 810

  269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
      |||||||   ::::::::::::   |||||||:::   :::|||:::||||:
  811 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 860

  286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
      ::|||||||   :::|||||||   :::|||:::   ||||||||||||:::
  861 TGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 910

  303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
      |||:::::::::|||||||:::|||   ::::::|||||||:::|||:::||
  911 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 960

  319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
      |:::::|||||||||      :::|||:::         ||||||   |
  961 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1010

  336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
      |||||      |||   |||      ||||||||||||   |||:::|||

 1011 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1060

  353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
      :::||||||:::|||   |||::::::|||:::||||||   ||||||||
 1061 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1110

  369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
      |   ::::::   ||||||:::||||||||||:::|||      :::   |
 1111 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1160

  386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
      ||:::||||||||||||:::::|||||||||:::::::::|||||||||
 1161 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1210

  403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
      |||:::   :::|||   ::::::      |||
 1211 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1243
```

```
seq_name: gb_htc:AK009473

seq_documentation_block:
LOCUS       AK009473     1345 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310022H17, full insert sequence.
ACCESSION   AK009473
VERSION     AK009473.1  GI:12844293
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310022H17.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1345)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1345)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1345)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1345)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1345)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
```

```
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES            Location/Qualifiers
     source         1. .1345
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /db_xref="MGD:MGI:1898391"
                    /db_xref="MGD:MGI:1914967"
                    /clone="2310022H17"
                    /sex="male"
                    /tissue_type="tongue"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="adult"
     CDS            52. .1239
                    /note="putative"
                    /codon_start=1
                    /protein_id="BAB26312.1"
                    /db_xref="GI:12844294"
                    /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                    SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                    FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                    GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                    FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                    YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                    AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                    AED"
BASE COUNT      395 a    288 c    283 g    379 t
ORIGIN


alignment_scores:
            Quality: 1138.50            Length:   411
              Ratio:  3.492               Gaps:     4
 Percent Similarity:  79.319   Percent Identity: 52.555

alignment_block:
 US-09-578-063-47 x AK009473   ..

 Align seg 1/1  to: AK009473  from: 1  to: 1345

       7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
         |||:::|||           :::   :::|||||||||||||:::::::
      34 AGGCACTGG............CACTACAAGATGTGGCTGCTATTAGTAAC 71

      23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
         :::: . :::           ::::::::::::::
```

```
  72 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 121

  39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
            :::|||||||||   |||||||:::|||:::::::|||   :::
 122 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 171

  56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
       |||||||||   ||||||||||||||:::|||||||||||||||||
 172 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 221

  72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
     ||||:::|||   |||||||||   |||       :::   ::::::   |
 222 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 271

  89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
     ||:::||||||||||:::   ||||||||||||||:::::::::||||:::
 272 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 321

 106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
     ||||||||||:::|||||||||||||||||||||:::|||||||||||||
 322 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 371

 122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
     |||||||:::||||||||:::|||||||||||||||||||:::|||||||
 372 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 421

 139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
     |||||:::           |||   |||:::   ||||||||||||||||||
 422 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 471

 156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
     :::|||||||||||:::::::|||||||||||:::|||:::|||||::
 472 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 521

 172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
     :||||||||||||||||||||||:::||||||||:::||||||||||||
 522 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 571

 189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
     |||||:::|||||||||||||||||||||   |||   ||||||:::|||
 572 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 621

 206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
     ||||||   ::::::|||||||||||||:::|||||||||||||:::::::
 622 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 671

 222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
     :||||||      |||:::   |||:::|||   ::::::::::|||   :
 672 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 721

 239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
     ::|||||||:::|||   |||:::        |||       :::||||||
 722 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 771

 253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
               |||          |||:::::::   :::||||||   :::||
 772 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 812

 269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
     |||||||   ::::::::::::   |||||||:::   :::|||:::|||:
 813 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 862

 286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
     ::|||||||   :::|||||||   :::|||:::   ||||||||||||:::
 863 TGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 912

 303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
     |||::::::::::||||||:::|||   :::::::||||||:::|||:::||
 913 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 962

 319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
     |:::::::||||||||      :::|||:::          ||||||   |
 963 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1012
```

```
     336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
         |||||      |||   |||      ||||||||||||   |||:::|||
    1013 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1062

     353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
         :::||||||:::|||   |||::::::|||:::||||||   ||||||||
    1063 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1112

     369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
         |   ::::::   ||||||:::||||||||||:::|||      :::   |
    1113 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1162

     386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
         ||:::|||||||||||||:::::::||||||||:::::::::|||||||||
    1163 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1212

     403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
         |||:::   :::|||   ::::::      |||
    1213 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1245
```

seq_name: gb_htc:AK009523

seq_documentation_block:

LOCUS AK009523 1346 bp mRNA HTC 05-JUL-2001

DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310028I09, full insert sequence.

ACCESSION AK009523

VERSION AK009523.1 GI:12844369

KEYWORDS CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2310028I09.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1346)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE 2 (bases 1 to 1346)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE 3 (bases 1 to 1346)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome research. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE 4 (bases 1 to 1346)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5 (bases 1 to 1346)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES Location/Qualifiers

source 1. .1346
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1895759"
/db_xref="MGD:MGI:1914967"
/clone="2310028I09"
/sex="male"
/tissue_type="tongue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"

CDS 53. .1240
/note="putative"
/codon_start=1
/protein_id="BAB26338.1"
/db_xref="GI:12844370"
/translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
AED"

BASE COUNT 395 a 289 c 283 g 379 t

ORIGIN

alignment_scores:

```
            Quality: 1138.50              Length:    411
              Ratio:  3.492                Gaps:      4
 Percent Similarity:  79.319   Percent Identity: 52.555
```

alignment_block:

US-09-578-063-47 x AK009523 ..

Align seg 1/1 to: AK009523 from: 1 to: 1346

```
      7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
        |||:::|||          :::   :::||||||||||||||:::::::
     35 AGGCACTGG.....CACTACAAGATGTGGCTGCTATTAGTAAC 72
```

```
 23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
    ::::   :::          ::::::::::::::
 73 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 122

 39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
           :::||||||||||   ||||||:::|||:::::::|||   :::
123 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 172

 56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
      |||||||||   |||||||||||||||:::|||||||||||||||||
173 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 222

 72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
    ||||:::|||   |||||||||   |||      :::   ::::::   |
223 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 272

 89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
    ||:::||||||||||:::   |||||||||||||||||:::::::::|||:::
273 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 322

106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
    |||||||||:::|||||||||||||||||||||||:::|||||||||||||||
323 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 372

122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
    |||||||:::|||||||||:::|||||||||||||||||||||:::|||||||
373 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 422

139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
    |||||:::          |||   |||:::   ||||||||||||||||||
423 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 472

156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
    :::|||||||||||||:::::::||||||||||||:::|||:::|||||||::
473 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 522

172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
    :|||||||||||||||||||||||||:::||||||||||:::||||||||||
523 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 572

189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
    |||||:::|||||||||||||||||||||||   |||  ||||||:::|||
573 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 622

206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
    ||||||   ::::::|||||||||||||:::|||||||||||||:::::::::
623 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 672

222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
    :||||||      |||:::   |||:::|||   ::::::::::|||   :
673 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 722

239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
    ::|||||:::|||   |||:::         |||      :::||||||
723 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 772

253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
           |||         |||::::::   :::|||||   :::||
773 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 813

269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
    |||||||   ::::::::::::   |||||||:::   :::|||:::|||:
814 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 863

286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
    ::||||||   :::||||||   :::|||:::   ||||||||||||:::
864 TGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 913

303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
    |||::::::::::|||||||:::|||   ::::::|||||||:::|||:::||
914 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 963

319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
    |:::::|||||||||      :::|||:::          ||||||   |
964 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1013

336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
    |||||      |||   |||      ||||||||||||   |||:::|||
1014 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1063

353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
    :::|||||:::|||   |||:::::::|||:::||||||   ||||||||
1064 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1113

369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
    |   ::::::   ||||||:::|||||||||:::|||      :::   |
1114 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1163

386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
    ||:::|||||||||||||:::::::||||||||:::::::::|||||||||
1164 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1213

403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
    |||:::   :::|||   ::::::      |||
1214 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1246
```

seq_name: gb_htc:AK009729

seq_documentation_block:

LOCUS AK009729 1346 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310040L03, full insert sequence.
ACCESSION AK009729
VERSION AK009729.1 GI:12844702
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2310040L03.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1346)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 1346)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (bases 1 to 1346)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome research. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (bases 1 to 1346)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 1346)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,

```
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES            Location/Qualifiers
     source         1. .1346
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /db_xref="MGD:MGI:1898390"
                    /db_xref="MGD:MGI:1914967"
                    /clone="2310040L03"
                    /sex="male"
                    /tissue_type="tongue"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="adult"
     CDS            53. .1240
                    /note="putative"
                    /codon_start=1
                    /protein_id="BAB26466.1"
                    /db_xref="GI:12844703"
                    /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                    SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                    FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                    GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                    FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                    YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                    AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                    AED"
BASE COUNT      395 a    288 c    283 g    380 t
ORIGIN


alignment_scores:
           Quality: 1138.50            Length:    411
             Ratio:  3.492              Gaps:      4
 Percent Similarity:  79.319   Percent Identity: 52.555

alignment_block:
 US-09-578-063-47 x AK009729   ..

 Align seg 1/1  to: AK009729  from: 1  to: 1346
```

```
  7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
    |||:::|||            :::   :::|||||||||||||:::::::
 35 AGGCACTGG............CACTACAAGATGTGGCTGCTATTAGTAAC 72

 23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
    ::::   :::           ::::::::::::::
 73 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 122

 39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
          :::||||||||||   ||||||:::|||:::::|||   :::
123 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 172

 56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
    |||||||||   ||||||||||||||:::||||||||||||||||||||
173 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 222

 72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
    ||||:::|||   |||||||||   |||      :::   ::::::   |
223 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 272

 89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
    ||:::|||||||||:::   |||||||||||||||:::::::::|||:::
273 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 322

106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
    |||||||||:::||||||||||||||||||||:::|||||||||||||||
323 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 372

122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
    |||||||:::|||||||||:::|||||||||||||||||||:::|||||||
373 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 422

139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
    ||||||:::          |||   |||:::   ||||||||||||||||||
423 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 472

156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
    :::|||||||||||:::::|||||||||||:::|||:::|||||||::
473 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 522

172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
    :|||||||||||||||||||||||:::|||||||:::||||||||||
523 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 572

189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
    |||||:::|||||||||||||||||||||   |||   ||||||:::|||
573 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 622

206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
    ||||||   ::::::|||||||||||:::|||||||||||||||::::::
623 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 672

222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
    :||||||      |||:::   |||:::|||   :::::::::|||   :
673 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 722

239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
    ::||||||:::|||   |||:::         |||      :::||||||
723 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 772

253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
          |||         |||::::::   :::|||||   :::||
773 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 813

269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
    |||||||   ::::::::::::   |||||||:::   :::|||:::|||:
814 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 863

286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
    ::|||||||   :::|||||||   :::|||:::   ||||||||||||:::
864 TGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 913

303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
```

```
          |||::::::::::|||||||:::|||   ::::::|||||||:::|||:::||
      914 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 963

      319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
          |::::::|||||||||      :::|||:::         ||||||   |
      964 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1013

      336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
          |||||      |||  ||      ||||||||||||   |||:::|||
     1014 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1063

      353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
          :::|||||||:::|||   |||:::::::|||:::||||||   ||||||||
     1064 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1113

      369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
          |   ::::::   ||||||:::||||||||||:::|||      :::   |
     1114 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1163

      386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
          ||:::|||||||||||||:::::|||||||||:::::::::|||||||||
     1164 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1213

      403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
          |||:::   :::|||   ::::::      |||
     1214 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1246

seq_name: gb_htc:AK010116

seq_documentation_block:
LOCUS       AK010116     1347 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310068O05, full insert sequence.
ACCESSION   AK010116
VERSION     AK010116.1  GI:12845334
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310068O05.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1347)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1347)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1347)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1347)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1347)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1347
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1895801"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310068O05"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              51. .1238
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26711.1"
                      /db_xref="GI:12845335"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFGV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      393 a    290 c    284 g    380 t
ORIGIN


alignment_scores:
            Quality: 1138.50             Length:    411
              Ratio:  3.492               Gaps:      4
 Percent Similarity:  79.319   Percent Identity: 52.555

alignment_block:
 US-09-578-063-47 x AK010116   ..
```

```
Align seg 1/1  to: AK010116  from: 1  to: 1347

      7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
        |||:::|||            :::   :::|||||||||||||:::::::
     33 AGGCACTGG...........CACTACAAGATGTGGCTGCTATTAGTAAC 70

     23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
        ::::   :::            :::::::::::::::
     71 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 120

     39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
           :::|||||||||   ||||||:::|||:::::::|||   :::
    121 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 170

     56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
        |||||||||   ||||||||||||||||:::||||||||||||||||||
    171 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 220

     72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
        ||||:::|||   |||||||||   |||      :::   ::::::   |
    221 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 270

     89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
        ||:::|||||||||:::   ||||||||||||||||:::::::::|||:::
    271 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 320

    106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
        |||||||||:::|||||||||||||||||||||||:::|||||||||||||
    321 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 370

    122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
        |||||||:::|||||||||:::|||||||||||||||||||:::|||||||
    371 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 420

    139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
        ||||||:::       |||   |||:::   ||||||||||||||||||
    421 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 470

    156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
        :::|||||||||||||:::::|||||||||||||:::|||:::|||||::
    471 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 520

    172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
        :|||||||||||||||||||||||:::|||||||||:::||||||||||
    521 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 570

    189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
        |||||:::|||||||||||||||||||||   |||   ||||||:::|||
    571 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 620

    206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
        ||||||   ::::::|||||||||||||:::|||||||||||||:::::::
    621 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 670

    222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
        :||||||      |||:::   |||:::|||  :::::::::|||   :
    671 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 720

    239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
        ::|||||||:::|||   |||:::         |||      :::||||||
    721 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 770

    253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
           |||         |||:::::::   :::||||||   :::||
    771 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 811

    269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
        |||||||   ::::::::::::   ||||||:::   :::|||:::|||:
    812 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 861

    286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
        ::|||||||   :::|||||||   :::|||:::   |||||||||||:::

    862 TGAGTCGCTTTGGTGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 911

    303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
        |||:::::::::|||||||:::|||   ::::::|||||||:::|||:::||
    912 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 961

    319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
        |:::::|||||||||      :::|||:::         ||||||   |
    962 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1011

    336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
        |||||      |||  |||      ||||||||||||   |||:::|||
   1012 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1061

    353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
        :::|||||::|||   |||:::::|||:::||||||   ||||||||
   1062 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1111

    369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
        |   ::::::   ||||||:::|||||||||:::|||      :::   |
   1112 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1161

    386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
        ||:::|||||||||||||:::::|||||||||:::::::::|||||||||
   1162 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1211

    403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
        |||:::   :::|||   ::::::      |||
   1212 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1244

seq_name: gb_htc:AK010058

seq_documentation_block:
LOCUS       AK010058     1348 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310066C03, full insert sequence.
ACCESSION   AK010058
VERSION     AK010058.1  GI:12845238
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310066C03.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1348)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1348)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1348)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1348)
```

```
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1348)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1348
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1895768"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310066C03"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             52. .1239
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26673.1"
                     /db_xref="GI:12845239"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      396 a    290 c    284 g    378 t
ORIGIN


alignment_scores:
            Quality: 1138.50            Length:    411
              Ratio:  3.492              Gaps:      4
 Percent Similarity:  79.319   Percent Identity: 52.555
```

```
alignment_block:
 US-09-578-063-47 x AK010058   ..

 Align seg 1/1  to: AK010058  from: 1  to: 1348

      7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
        |||:::|||           :::   :::|||||||||||||:::::::
     34 AGGCACTGG...........CACTACAAGATGTGGCTGCTATTAGTAAC 71

     23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
        ::::   :::          ::::::::::::::
     72 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 121

     39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
              :::||||||||||   ||||||:::|||:::::::|||   :::
    122 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 171

     56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
         |||||||||   ||||||||||||||:::||||||||||||||||||
    172 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 221

     72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
        ||||:::|||   |||||||||   |||       :::   ::::::   |
    222 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 271

     89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
        ||:::|||||||||:::   ||||||||||||||||:::::::::|||:::
    272 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 321

    106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
        |||||||||:::||||||||||||||||||:::|||||||||||||||
    322 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 371

    122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
        |||||||:::|||||||||:::||||||||||||||||||||:::|||||||
    372 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 421

    139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
        |||||:::          |||   |||:::   ||||||||||||||||||
    422 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 471

    156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
        :::|||||||||||:::::||||||||||||:::|||:::|||||||::
    472 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 521

    172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
        :||||||||||||||||||||:::||||||||::||||||||||
    522 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 571

    189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
        |||||:::|||||||||||||||||||||   |||   ||||||:::|||
    572 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 621

    206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
        ||||||   ::::::||||||||||::||||||||||||||:::::::
    622 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 671

    222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
        :||||||      |||:::   |||:::|||   ::::::::::|||   :
    672 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 721

    239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
        ::|||||:::|||   |||:::         |||       ::|||||||
    722 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 771

    253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
                |||         |||::::::   :::||||||   :::||
    772 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 812

    269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
        |||||||   ::::::::::::   ||||||:::   :::|||:::|||:
    813 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 862
```

```
     286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrServal 302
         ::|||||||   :::|||||||   :::|||:::   |||||||||||||:::
     863 TGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 912

     303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
         |||::::::::::|||||||:::|||   :::::::|||||||:::|||:::||
     913 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 962

     319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
         |:::::::|||||||||      :::|||:::          ||||||   |
     963 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1012

     336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
         |||||      |||   |||      ||||||||||||   |||:::|||
    1013 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1062

     353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
         :::||||||:::|||   |||:::::::|||:::|||||||   ||||||||
    1063 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1112

     369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
         |   ::::::   ||||||:::||||||||||:::|||      :::   |
    1113 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1162

     386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
         ||:::|||||||||||||:::::|||||||||:::::::::|||||||||
    1163 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1212

     403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
         |||:::   :::|||   ::::::      |||
    1213 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1245
```

seq_name: gb_htc:AK009300

seq_documentation_block:

```
LOCUS       AK009300     1349 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310011G18, full insert sequence.
ACCESSION   AK009300
VERSION     AK009300.1  GI:12844011
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310011G18.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1349)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1349)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1349)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
   MEDLINE  20530913
    PUBMED  11076861
REFERENCE   4  (bases 1 to 1349)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1349)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1349
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1895751"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310011G18"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              53. .1240
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26201.1"
                      /db_xref="GI:12844012"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFGV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      395 a    290 c    284 g    380 t
ORIGIN
```

alignment_scores:

```
          Quality: 1138.50            Length:    411
            Ratio:  3.492              Gaps:      4
 Percent Similarity:  79.319   Percent Identity: 52.555

alignment_block:
 US-09-578-063-47 x AK009300   ..

 Align seg 1/1  to: AK009300  from: 1  to: 1349

       7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
         |||:::|||           :::   :::||||||||||||:::::::
      35 AGGCACTGG...........CACTACAAGATGTGGCTGCTATTAGTAAC 72

      23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
         ::::   :::          ::::::::::::::
      73 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 122

      39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
               :::||||||||||   ||||||:::|||:::::::|||   :::
     123 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 172

      56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
            |||||||||   |||||||||||||||:::|||||||||||||||||
     173 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 222

      72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
         ||||:::|||   |||||||||   |||       :::   ::::::   |
     223 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 272

      89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
         ||:::||||||||||:::   |||||||||||||||||:::::::::|||:::
     273 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 322

     106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
         |||||||||:::|||||||||||||||||||||||::::|||||||||||||
     323 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 372

     122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
         |||||||:::|||||||||:::|||||||||||||||||||:::|||||||
     373 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 422

     139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
         |||||:::          |||   |||:::   ||||||||||||||||||
     423 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 472

     156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
         :::|||||||||||||:::::|||||||||||:::|||:::||||||::
     473 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 522

     172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
         :|||||||||||||||||||||||:::||||||||:::||||||||||||
     523 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 572

     189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
         |||||:::|||||||||||||||||||||   |||   ||||||:::|||
     573 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 622

     206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
         ||||||   ::::::|||||||||||||:::|||||||||||:::::::::
     623 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 672

     222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
         :||||||      |||:::   |||:::|||   :::::::::|||   :
     673 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 722

     239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
         ::||||||:::|||   |||:::         |||     :::||||||
     723 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 772

     253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
               |||         |||:::::   :::||||||   :::||
     773 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 813

     269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
         |||||||   ::::::::::::   |||||||:::   :::|||:::||||:
     814 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 863

     286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
         ::||||||   :::||||||   :::|||:::   ||||||||||||:::
     864 TGAGTCGCTTTGGTGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 913

     303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
         |||:::::::::||||||:::|||   ::::::||||||:::|||:::||
     914 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 963

     319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
         |:::::::|||||||||      :::|||:::          ||||||   |
     964 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1013

     336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
         |||||      |||   |||      ||||||||||||   |||:::|||
    1014 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1063

     353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
         :::||||||:::|||   |||:::::::|||:::||||||   ||||||||
    1064 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1113

     369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
         |   ::::::   ||||||:::|||||||:::|||      :::   |
    1114 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1163

     386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
         ||:::|||||||||||:::::::||||||||:::::::::|||||||||
    1164 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1213

     403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
         |||:::   :::|||   ::::::      |||
    1214 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1246

seq_name: gb_htc:AK009474

seq_documentation_block:
LOCUS       AK009474     1349 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310022I03, full insert sequence.
ACCESSION   AK009474
VERSION     AK009474.1  GI:12844295
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310022I03.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1349)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1349)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1349)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
```

```
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1349)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1349)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1349
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1898389"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310022I03"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53. .1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26313.1"
                     /db_xref="GI:12844296"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      397 a    290 c    284 g    378 t
ORIGIN
```

```
alignment_scores:
            Quality: 1138.50              Length:    411
              Ratio:  3.492                Gaps:      4
 Percent Similarity:  79.319   Percent Identity: 52.555

alignment_block:
 US-09-578-063-47 x AK009474   ..

 Align seg 1/1  to: AK009474  from: 1  to: 1349

       7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
         |||:::|||             :::   :::|||||||||||||:::::::
      35 AGGCACTGG.............CACTACAAGATGTGGCTGCTATTAGTAAC 72

      23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
         ::::   :::            ::::::::::::::
      73 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 122

      39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
            :::||||||||||  ||||||:::|||:::::::|||   :::
     123 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 172

      56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
            |||||||||   ||||||||||||||||:::||||||||||||||||||
     173 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 222

      72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
         ||||:::|||   |||||||||   |||       :::   ::::::   |
     223 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 272

      89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
         ||:::|||||||||:::   ||||||||||||||:::::::::|||:::
     273 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 322

     106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
         |||||||||:::|||||||||||||||||||||:::|||||||||||||||
     323 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 372

     122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
         |||||||:::|||||||:::|||||||||||||||||||:::|||||||
     373 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 422

     139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
         |||||:::       |||   |||:::  ||||||||||||||||||
     423 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 472

     156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
         :::|||||||||||||:::::::|||||||||||||:::|||:::|||||::
     473 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 522

     172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
         :|||||||||||||||||||||||:::|||||||||::|||||||||||
     523 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 572

     189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
         |||||:::|||||||||||||||||||||||   |||   ||||||:::|||
     573 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 622

     206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
         ||||||   ::::::|||||||||||||:::|||||||||||||:::::::
     623 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 672

     222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
         :||||||      |||:::  |||:::|||   ::::::::::|||   :
     673 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 722

     239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
         ::||||||:::|||   |||:::          |||       :::||||||
     723 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 772

     253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
```

```
                |||          |||::::::   :::||||||   :::||
    773 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 813

    269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
        |||||||   ::::::::::::   ||||||:::   :::|||:::|||:
    814 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 863

    286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
        ::||||||   :::||||||   :::|||:::   ||||||||||||:::
    864 TGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 913

    303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
        |||::::::::::||||||:::|||   ::::::||||||:::|||:::||
    914 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 963

    319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
        |:::::|||||||||      :::|||:::          ||||||   |
    964 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1013

    336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
        |||||      |||   |||      ||||||||||||   |||:::|||
   1014 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1063

    353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
        :::||||||:::|||   |||:::::|||:::||||||   ||||||||
   1064 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1113

    369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
        |   ::::::   ||||||:::||||||||||:::|||      :::   |
   1114 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1163

    386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
        ||:::||||||||||||:::::|||||||||:::::::::||||||||||
   1164 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1213

    403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
        |||:::   :::|||   ::::::      |||
   1214 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1246
```

seq_name: gb_htc:AK009546

seq_documentation_block:

```
LOCUS       AK009546     1349 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310030L19, full insert sequence.
ACCESSION   AK009546
VERSION     AK009546.1  GI:12844406
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310030L19.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1349)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1349)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1349)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1349)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1349)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES            Location/Qualifiers
     source         1. .1349
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /db_xref="MGD:MGI:1893997"
                    /db_xref="MGD:MGI:1914967"
                    /clone="2310030L19"
                    /sex="male"
                    /tissue_type="tongue"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="adult"
     CDS            56. .1243
                    /note="putative"
                    /codon_start=1
                    /protein_id="BAB26352.1"
                    /db_xref="GI:12844407"
                    /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                    SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                    FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                    GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                    FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                    YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                    AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
```

```
                    AED"
BASE COUNT      396 a    290 c    284 g    379 t
ORIGIN

alignment_scores:
            Quality: 1138.50              Length:    411
              Ratio:  3.492                Gaps:      4
 Percent Similarity:  79.319   Percent Identity: 52.555

alignment_block:
 US-09-578-063-47 x AK009546   ..

 Align seg 1/1  to: AK009546  from: 1  to: 1349

       7 ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23
         |||:::|||            :::   :::|||||||||||||:::::::
      38 AGGCACTGG............CACTACAAGATGTGGCTGCTATTAGTAAC 75

      23 lAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
         ::::   :::           ::::::::::::::
      76 AAGTGTGCTATCTGCATTTGGAGGTGCACATGGCCTATTTGGAAAACTGG 125

      39 ysAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
               :::|||||||||   ||||||:::|||:::::::|||   :::
     126 GTCCCAAAAACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTAC 175

      56 GlnGlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIl 72
             |||||||||   ||||||||||||||:::|||||||||||||||||
     176 TGGGGATATCCAAGTGAGGAATATGAAGTTGTTACTGAAGATGGCTACAT 225

      72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
         ||||:::|||   |||||||||   |||       :::   ::::::   |
     226 TCTGGGGGTCTATAGAATTCCTTATGGGAAGAAAAATTCTGAGAATATCG 275

      89 lySerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
         ||:::|||||||||:::   ||||||||||||||:::::::::|||:::
     276 GCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTGCATCAGCCACA 325

     106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
         |||||||||:::|||||||||||||||||||:::|||||||||||||||
     326 AACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAGA 375

     122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
         |||||||:::|||||||||:::|||||||||||||||||||:::|||||||
     376 TGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCC 425

     139 rgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
         |||||:::          |||  |||:::   ||||||||||||||||||
     426 GGAAAAATGTATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGC 475

     156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172
         :::|||||||||||||:::::::|||||||||||||:::|||:::|||||::
     476 TTTGATGAAATGGCTAAATATGACCTTCCAGCCACCATAGACTTCATTGT 525

     172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyT 189
         :|||||||||||||||||||||||:::||||||||||:::||||||||||
     526 ACAGAAAACTGGACAAGAGAAGATACACTATGTTGGTCACTCTCAGGGCA 575

     189 hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
         |||||:::|||||||||||||||||||||   |||   ||||||:::|||
     576 CCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTCTGGCTAAAAAA 625

     206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaLy 222
         ||||||   ::::::|||||||||||||:::|||||||||||||:::::::
     626 ATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAGA 675

     222 sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyL 239
         :||||||      |||:::   |||:::|||   ::::::::::||| :
     676 AAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGA 725

     239 euPheGlyLysLysGluPheLeu.........TyrGlnThrArgPheLeu 252
         ::|||||||:::|||   |||:::         |||      :::|||||||

     726 TATTTGGTAACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTT 775

     253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCy 269
                |||         |||:::::::   :::|||||||   :::||
     776 GGTACGGAAGTG.........TGCTCACGGGAGCTGCTAGATCTTCTCTG 816

     269 sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
         |||||||   ::::::::::::   |||||||:::   :::|||:::||||:
     817 CAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGAAAAACTTAAATG 866

     286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
         ::||||||   :::||||||   :::|||:::   |||||||||||:::
     867 TGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTACT 916

     303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
         |||::::::::::||||||:::|||   ::::::||||||:::|||:::||
     917 CAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGC 966

     319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
         |:::::||||||||      :::|||:::         ||||||   |
     967 CTATAACTGGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAA 1016

     336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
         |||||      |||   |||      ||||||||||||   |||:::|||
    1017 CGCCTCCCTACTATGATGTGTCAGCCATGACCGTGCCAATTGCAGTGTGG 1066

     353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369
         :::|||||||:::|||   |||::::::|||:::||||||   ||||||||
    1067 AACGGTGGCCATGACATCCTGGCTGATCCCCAAGATGTCGCAATGCTGCT 1116

     369 uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
         |   ::::::   ||||||:::|||||||||:::|||      :::   |
    1117 TCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTCTTCCCTACAATC 1166

     386 isValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
         ||:::|||||||||||:::::::||||||||:::::::::||||||||||
    1167 ACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATGAG 1216

     403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
         |||:::   :::|||   ::::::      |||
    1217 ATAGTTACCATGATGGCAGAAGACTAACAGAAT 1249
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 30, 2002, 22:09:19 ; Search time 67.96 Seconds
(without alignments)
461.051 Million cell updates/sec

Title: US-09-578-063-47
Perfect score: 2247
Sequence: 1 MLETLSRQWIVSHRMEMWLL.....IHLMQQEETNLSQGRCEAVL 423

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2247	100.0	423	22	AAB66065	Human TANGO 294.
2	2076	92.4	390	22	AAB66067	Human TANGO 294 ma
3	1289	57.4	399	22	AAB90783	Human shear stress
4	1289	57.4	399	22	AAB66061	Human lysosomal ac
5	1166	51.9	398	14	AAR37302	RGL precursor. Or
6	1165	51.8	221	22	AAB66068	Human TANGO 294 ex
7	1156.5	51.5	395	6	AAP50322	Rat lingual lipase
8	1141.5	50.8	379	15	AAR56870	Canine gastric lip
9	1141.5	50.8	379	17	AAW09382	Dog gastric lipase
10	1141.5	50.8	380	15	AAR56871	Canine gastric lip
11	1119	49.8	398	7	AAP60724	Sequence of pregas
12	1119	49.8	398	7	AAP60658	Sequence of human
13	1119	49.8	398	17	AAW09383	Human gastric lipa
14	1119	49.8	398	22	AAB66086	Human lipase prote
15	783	34.8	144	22	AAB66070	Human TANGO 294 cy
16	691	30.8	233	22	AAB61608	Human protein HP03
17	514.5	22.9	410	21	AAY32309	Soybean acid triac
18	502	22.3	410	21	AAY32307	Rice acid triacylg
19	465	20.7	405	21	AAY32310	Soybean acid triac
20	398	17.7	286	21	AAY32304	Corn acid triacylg
21	262	11.7	157	21	AAY32303	Catalpa acid triac
22	171	7.6	33	22	AAB66066	Human TANGO 294 si
23	160.5	7.1	125	21	AAY32305	Corn acid triacylg
24	150	6.7	143	21	AAY32301	Corn acid triacylg
25	138	6.1	90	21	AAY32306	Corn acid triacylg
26	138	6.1	116	21	AAY32302	Catalpa acid triac
27	128	5.7	25	22	AAB66069	Human TANGO 294 tr
28	105	4.7	430	18	AAW22234	C. elegans origin
29	105	4.7	430	18	AAW14140	Caenorhabditis ele
30	102	4.5	346	18	AAW23088	Pyrodictium sp. es
31	101	4.5	308	17	AAR91905	Pseudomonas wiscon
32	100.5	4.5	286	17	AAR91903	Pseudomonas wiscon
33	100.5	4.5	286	18	AAW22309	P. wisconsiensis m
34	99	4.4	678	22	AAG92271	C glutamicum prote
35	98.5	4.4	388	15	AAR72124	Acetobacter pasteu
36	98	4.4	471	20	AAY00878	Sucrose fatty acid
37	98	4.4	806	20	AAY21797	Alteromonas lipase
38	98	4.4	809	20	AAY21796	Alteromonas lipase
39	96.5	4.3	1010	21	AAB38012	Human secreted pro
40	95	4.2	336	22	AAM39139	Human polypeptide
41	94.5	4.2	340	21	AAG20361	Arabidopsis thalia
42	94.5	4.2	392	21	AAG20360	Arabidopsis thalia
43	94.5	4.2	1280	22	AAB64943	Human secreted pro
44	93.5	4.2	484	21	AAG45186	Arabidopsis thalia
45	93	4.1	637	22	AAG90967	C glutamicum prote

ALIGNMENTS

RESULT 1
AAB66065
ID AAB66065 standard; Protein; 423 AA.
XX
AC AAB66065;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294.
XX
KW TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
DR N-PSDB; AAF45131, AAF45132.
XX

PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease -
XX
PS Claim 8; Fig 6; 359pp; English.
XX
CC The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other
CC neurological and cerebrovascular disorders. The CNS disorders include
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC autonomic function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder.
XX
SQ Sequence 423 AA;

Query Match 100.0%; Score 2247; DB 22; Length 423;
Best Local Similarity 100.0%; Pred. No. 4.5e-220;
Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MLETLSRQWIVSHRMEMWLLILVAYMFQRNVNSVHMPTKAVDPEAFMNISEIIQHQGYPC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mletlsrqwivshrmemwllilvaymfqrnvnsvhmptkavdpeafmniseiiqhqgypc 60

Qy      61 EEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFIL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 eeyevatedgyilsvnriprglvqpkktgsrpvvllqhglvggasnwisnlpnnslgfil 120

Qy     121 ADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 adagfdvwmgnsrgnawsrkhktlsidqdefwafsydemarfdlpavinfilqktgqeki 180

Qy     181 YYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLF 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 yyvgysqgttmgfiafstmpelaqkikmyfalapiatvkhakspgtkflllpdmmikglf 240

Qy     241 GKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 gkkeflyqtrflrqlviylcgqvildqicsnimlllggfntnnmnmsrasvyaahtlagt 300

Qy     301 SVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLS 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 svqnilhwsqavnsgelrafdwgsetknlekcnqptpvryrvrdmtvptamwtggqdwls 360

Qy     361 NPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQEETNLSQGRCE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 npedvkmllsevtnliyhknipewahvdfiwgldaphrmyneiihlmqqeetnlsqgrce 420

Qy     421 AVL 423
          |||
Db     421 avl 423
```

RESULT 2
AAB66067
ID AAB66067 standard; Protein; 390 AA.
XX
AC AAB66067;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294 mature protein.
XX
KW TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
XX
PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease -
XX
PS Claim 8; Pages 324-325; 359pp; English.
XX
CC The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other
CC neurological and cerebrovascular disorders. The CNS disorders include
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC autonomic function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder.
XX
SQ Sequence 390 AA;

Query Match 92.4%; Score 2076; DB 22; Length 390;
Best Local Similarity 100.0%; Pred. No. 1e-202;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      34 VHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPV 93
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 vhmptkavdpeafmniseiiqhqgypceeyevatedgyilsvnriprglvqpkktgsrpv 60

Qy      94 VLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWA 153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 vllqhglvggasnwisnlpnnslgfiladagfdvwmgnsrgnawsrkhktlsidqdefwa 120

Qy     154 FSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALA 213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 fsydemarfdlpavinfilqktgqekiyyvgysqgttmgfiafstmpelaqkikmyfala 180

Qy     214 PIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIM 273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 piatvkhakspgtkfllldmmikglfgkkeflyqtrflrqlviylcgqvildqicsnim 240

Qy     274 LLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCN 333
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 lllggfntnnmnmsrasvyaahtlagtsvqnilhwsqavnsgelrafdwgsetknlekcn 300

Qy     334 QPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGL 393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 qptpvryrvrdmtvptamwtggqdwlsnpedvkmllsevtnliyhknipewahvdfiwgl 360

Qy     394 DAPHRMYNEIIHLMQQEETNLSQGRCEAVL 423
           |||||||||||||||||||||||||||||
```

```
Db    361 daphrmyneiihlmqqeetnlsqgrceavl 390


RESULT   3
AAB90783
ID   AAB90783 standard; Protein; 399 AA.
XX
AC   AAB90783;
XX
DT   15-JUN-2001  (first entry)
XX
DE   Human shear stress-response protein SEQ ID NO: 66.
XX
KW   Human; shear stress-response protein; vascular disease;
KW   arteriosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO200125427-A1.
XX
PD   12-APR-2001.
XX
PF   02-OCT-2000; 2000WO-JP06840.
XX
PR   01-OCT-1999;   99JP-0280976.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (NOJI/) NOJIMA H.
XX
PI   Nojima H,  Yoshisue H,  Obayashi M,  Ota T,  Kawabata A,  Sakurada K;
PI   Kuga T,  Sekine S,  Nakamura Y,  Sugano S;
XX
DR   WPI; 2001-266308/27.
DR   N-PSDB; AAH02906.
XX
PT   DNA sequences, proteins encoded by them and antibodies against them
PT   useful in diagnosis and treatment of vascular disease caused by
PT   arteriosclerosis -
XX
PS   Claim 60; Page 402-404; 678pp; Japanese.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human shear stress response proteins. These are useful in the
CC   diagnosis, treatment and screening of vascular diseases caused by
CC   arteriosclerosis, including heart failure, post-PTCA restenosis and
CC   hypertension.
XX
SQ   Sequence   399 AA;


  Query Match             57.4%;  Score 1289;  DB 22;  Length 399;
  Best Local Similarity   63.5%;  Pred. No. 1.5e-122;
  Matches  235;  Conservative   55;  Mismatches   80;  Indels    0;  Gaps    0;

Qy      40 AVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLLQHG 99
           |||||  ||:|||| : |:| ||| | ||||||| :|||| |       | :||| ||||
Db      28 avdpetnmnvseiisywgfpseeylvetedgyilclnriphgrknhsdkgpkpvvflqhg 87

Qy     100 LVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEM 159
           |:  :|||::|| |:|||||||||||||||||||| |||||||||: ||||||||||||
Db      88 lladssnwvtnlansslgfiladagfdvwmgnsrgntwsrkhktlsvsqdefwafsydem 147

Qy     160 ARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVK 219
           |::|||| ||||| |||||::||||:|||||:|||||| :||||::|||:||| |:|:|
Db     148 akydlpasinfilnktgqeqvyyvghsqgttigfiafsqipelakrikmffalgpvasva 207

Qy     220 HAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGF 279
              ||  |   ||| :|| ||| |||| |: ||: |  ::|  ||| ::| |:  || ||
Db     208 fctspmaklgrlpdhlikdlfgdkeflpqsaflkwlgthvcthvilkelcgnlcfllcgf 267

Qy     280 NTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVR 339
           |  |:|||  ||  |: |||||||:|||||||   : :||||||  ||    ||  |
Db     268 nernlnmsrvdvytthspagtsvqnmlhwsqavkfqkfqafdwgssaknyfhynqsyppt 327

Qy     340 YRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHRM 399
           | |:|| ||||:|:|| |||::  || :||:::|||::|::|||| |:||||||||| |:
Db     328 ynvkdmlvptavwsgghdwladvydvnilltqitnlvfhesipewehldfiwgldapwrl 387

Qy     400 YNEIIHLMQQ 409
           ||:||:||::
Db     388 ynkiinlmrk 397


RESULT   4
AAB66061
ID   AAB66061 standard; Protein; 399 AA.
XX
AC   AAB66061;
XX
DT   30-MAR-2001  (first entry)
XX
DE   Human lysosomal acid lipase protein.
XX
KW   Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
KW   central nervous system; focal brain disorder; bipolar affective disorder;
KW   global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW   senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW   Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW   neuropsychiatric; psychoactive substance use; anxiety.
XX
OS   Homo sapiens.
XX
PN   WO200077239-A2.
XX
PD   21-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14858.
XX
PR   14-JUN-1999;   99US-0333159.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   McCarthy SA,  Fraser CC,  Sharp JD,  Barnes TM;
XX
DR   WPI; 2001-032313/04.
XX
PT   TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT   screening assays and diagnostic assays and for the treatment of
PT   neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT   disease -
XX
PS   Claim 8; Fig 6; 359pp; English.
XX
CC   The present invention relates to TANGO or INTERCEPT proteins and coding
CC   sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC   AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC   sequences are useful for the treatment of neurological disorders such as
CC   central nervous system (CNS) disorders, CNS-related disorders, focal
CC   brain disorders, global-diffuse cerebral disorders and other
CC   neurological and cerebrovascular disorders. The CNS disorders include
CC   Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC   lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC   autonomic function disorders such as hypertension and sleep disorders,
CC   neuropsychiatric disorders, psychoactive substance use disorders,
CC   anxiety, and bipolar affective disorder. The present sequence is a
CC   sequence used in a sequence homology alignment with the TANGO/INTERCEPT
CC   sequences of the present invention.
XX
SQ   Sequence   399 AA;


  Query Match             57.4%;  Score 1289;  DB 22;  Length 399;
  Best Local Similarity   63.5%;  Pred. No. 1.5e-122;
  Matches  235;  Conservative   55;  Mismatches   80;  Indels    0;  Gaps    0;

Qy      40 AVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLLQHG 99
```

```
          |||||  ||:|||| : |:| ||| | ||||||| :|||| |       | :||| ||||
Db     28 avdpetnmnvseiisywgfpseeylvetedgyilclnriphgrknhsdkgpkpvvflqhg 87

Qy    100 LVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEM 159
          |:  :|||::|| |:|||||||||||||||||||| |||||||||: ||||||||||||
Db     88 lladssnwvtnlansslgfiladagfdvwmgnsrgntwsrkhktlsvsqdefwafsydem 147

Qy    160 ARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVK 219
          |::|||| |||||  |||||::|||||:|||||:|||||| :||||::|||:||| |:|:|
Db    148 akydlpasinfilnktgqeqvyyvghsqgttigfiafsqipelakrikmffalgpvasva 207

Qy    220 HAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGF 279
             ||  |   ||| :|| ||| |||| |: ||: |  ::|  ||| ::| |:  || ||
Db    208 fctspmaklgrlpdhlikdlfgdkeflpqsaflkwlgthvcthvilkelcgnlcfllcgf 267

Qy    280 NTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVR 339
          |  |:|||||  ||  |: |||||||:|||||||   : :||||||  ||    ||  |
Db    268 nernlnmsrvdvytthspagtsvqnmlhwsqavkfqkfqafdwgssaknyfhynqsyppt 327

Qy    340 YRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHRM 399
          | |:|| ||||:|:|| |||::  || :||:::|||::|::|||| |:|||||||| |:
Db    328 ynvkdmlvptavwsgghdwladvydvnilltqitnlvfhesipewehldfiwgldapwrl 387

Qy    400 YNEIIHLMQQ 409
          ||:||:||::
Db    388 ynkiinlmrk 397
```

RESULT 5
AAR37302
ID AAR37302 standard; Protein; 398 AA.
XX
AC AAR37302;
XX
DT 20-SEP-1993 (first entry)
XX
DE RGL precursor.
XX
KW Rabbit gastric lipase; RGL; pRGLN2.1; fat; bioconversion;
KW hydrolysis; transesterification.
XX
OS Oryctolagus cuniculus.
XX
FH Key Location/Qualifiers
FT Protein 23..398
FT /note= "claim 1; page 10-11"
XX
PN EP542629-A.
XX
PD 19-MAY-1993.
XX
PF 12-NOV-1992; 92EP-0403055.
XX
PR 13-NOV-1991; 91FR-0013948.
XX
PA (LJOU) INST RECH JOUVEINAL.
XX
PI Benicourt C, Blanchard C, Junien J;
XX
DR WPI; 1993-161080/20.
DR N-PSDB; AAQ42310.
XX
PT Rabbit gastric lipase, its precursor and their DNA - useful for
PT treating conditions linked to gastric lipase deficiency, such as
PT mucoviscidiosis and pancreatic exocrine insufficiency
XX
PS Claim 1; Fig 7; 31pp; French.
XX
CC RGL, opt. used with other lipases, are useful therapeutically
CC (1) to facilitate absorption of ingested fats in patients deficient
CC in endogenous gastic lipase and (2) to treat disorders caused by
CC inadequate lipase prodn. esp. mucoviscidiosis or pancreatic exocrine
CC deficiency. It can also be used (partic. when immobilised) for
CC enzymatic bioconversion, e.g. hydrolysis or transesterification.
XX
SQ Sequence 398 AA;

```
  Query Match             51.9%;  Score 1166;  DB 14;  Length 398;
  Best Local Similarity   54.6%;  Pred. No. 4.9e-110;
  Matches  219;  Conservative   70;  Mismatches  102;  Indels   10;  Gaps    3;

Qy     17 MWLLILVAYMFQRNVNSVH------MPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDG 70
          ||:| :|| :    : : |       ||   :||  ||||::| : ||| |:||| ||||
Db      1 mwvlfmvaallsa-lgtthglfgksapt---npevnmnisqmisywgypsekyevvtedg 56

Qy     71 YILSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMG 130
          ||| ||||| |       | |||| |||||:  ||||||||||||| |||||||: ||:|
Db     57 yilevnripygkknsgnrgqrpvvflqhgllasasnwisnlpnnslafiladagygvwlg 116

Qy    131 NSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTT 190
          ||||| |||::   | |  ||||||:||||::|||| |:||:::|||||::|||:|||||
Db    117 nsrgntwsrrnlyyspdsvefwafsfdemakydlpatidfivketgqeklhyvghsqgtt 176

Qy    191 MGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTR 250
          :|||||| |:||::|| ::||||:||||: ||   |   :|  | | :|| | |
Db    177 igfiafstnpklaeriktfyalapvatvkytkslvnklrfipptmfkiifgdkifyphnf 236

Qy    251 FLRQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQ 310
          | : |   :| :  |: |||| : :: ||:: |:|||  || :|  |||||||:|||:|
Db    237 fdqflatqvcsretlnvicsnalfiicgfdsanlnmsrldvyvshnpagtsvqnmlhwtq 296

Qy    311 AVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLS 370
          || ||  :||:|||  :|:  |||||  | |  | || |:|:|| |||::|:|| :||
Db    297 avksgnfqafnwgspaqnvvhfnqptppyynvtamnvpiavwsggndwladpqdvdlllp 356

Qy    371 EVTNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQEE 411
          :::|||||| |  : |:|||| ::||  :||||| :| :::
Db    357 klsnliyhkeilpynhldfiwamnapqevyneiismmakdk 397
```

RESULT 6
AAB66068
ID AAB66068 standard; Protein; 221 AA.
XX
AC AAB66068;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294 extracellular domain.
XX
KW TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
XX
PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of

PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease -
XX
PS Claim 8; Page 326; 359pp; English.
XX
CC The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other
CC neurological and cerebrovascular disorders. The CNS disorders include
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC autonomic function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder.
XX
SQ Sequence 221 AA;

Query Match 51.8%; Score 1165; DB 22; Length 221;
Best Local Similarity 100.0%; Pred. No. 2.5e-110;
Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      34 VHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPV 93
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 vhmptkavdpeafmniseiiqhqgypceeyevatedgyilsvnriprglvqpkktgsrpv 60

Qy      94 VLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWA 153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 vllqhglvggasnwisnlpnnslgfiladagfdvwmgnsrgnawsrkhktlsidqdefwa 120

Qy     154 FSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALA 213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 fsydemarfdlpavinfilqktgqekiyyvgysqgttmgfiafstmpelaqkikmyfala 180

Qy     214 PIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ 254
           |||||||||||||||||||||||||||||||||||||||||
Db     181 piatvkhakspgtkflllpdmmikglfgkkeflyqtrflrq 221
```

RESULT 7
AAP50322
ID AAP50322 standard; protein; 395 AA.
XX
AC AAP50322;
XX
DT 17-JAN-1992 (first entry)
XX
DE Rat lingual lipase protein.
XX
KW Lingual lipase; enzyme; EC-3.1.1.3; ss.
XX
OS Rattus rattus.
XX
PN GB2142337-A.
XX
PD 16-JAN-1985.
XX
PF 29-JUN-1984; 84GB-0016581.
XX
PR 01-JUL-1983; 83GB-0017989.
PR 05-SEP-1983; 83GB-0023759.
XX
PA (CELL-) CELLTECH LTD.
XX
PI Carey NH, Williamson R;
XX
DR WPI; 1985-014450/03.
DR N-PSDB; AAN50385.
XX
PT New lingual lipase protein for treatment of lipase deficiency - also
PT new pre-lingual lipase protein and related products
XX
PS Disclosure; Fig 5; 15pp; English.
XX
CC This protein may be expressed in a transformant host organism
CC and may be used for the treatment of lipase deficiency.
XX
SQ Sequence 395 AA;

Query Match 51.5%; Score 1156.5; DB 6; Length 395;
Best Local Similarity 54.7%; Pred. No. 4.5e-109;
Matches 216; Conservative 69; Mismatches 109; Indels 1; Gaps 1;

```
Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           |||||:: : :      : :       :||| ||||::| : ||||:|||| ||||||| |
Db       1 mwlllitsvistfggahglfgklgpgnpeanmnisqmitywgypcqeyevvtedgyilgv 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |   :  | |||| |||||:  |:|||:||||||| |:||||||:|||:||||||
Db      61 yriphgknnsenigkrpvvylqhgliasatnwianlpnnslafmladagydvwlgnsrgn 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           ||||:   | |  ||||||:||||::|||| ||||:||||||||:|||:|| ||:||||
Db     121 twsrknvyyspdsvefwafsfdemakydlpatinfivqktgqekihyvghsqrttigfia 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQL 255
           ||| | ||:||| ::||||:||||: :||  |   :|  : | :|||| ||  | |   |
Db     181 fstnptlakkiktfyalapvatvkytqsplkkisfiptflfklmfgkkmflphtyfddfl 240

Qy     256 VIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSG 315
           : :| : :|| :||| : :  ||:  |:|:||  ||  |  ||||||: |||:| | ||
Db     241 gtevcsrevldllcsntlfifcgfdkknlnvsrfdvylghnpagtsvqdflhwaqlvrsg 300

Qy     316 ELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNL 375
           : :||:||| ::|:   || ||  | |  |||| |:| || | |::|:|| ||| :::||
Db     301 kfqafnwgspsqnmlhynqktppeydvsamtvpvavwnggndiladpqdvamllpklsnl 360

Qy     376 IYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ::|| |  : |:|||| :|||  :|||:| :| ::
Db     361 lfhkeilaynhldfiwamdapqevynemismmaed 395
```

RESULT 8
AAR56870
ID AAR56870 standard; Protein; 379 AA.
XX
AC AAR56870;
XX
DT 20-FEB-1995 (first entry)
XX
DE Canine gastric lipase.
XX
KW Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
KW bioconversion; exocrine pancreatic insufficiency.
XX
OS Canis familiaris.
XX
PN WO9413816-A.
XX
PD 23-JUN-1994.
XX
PF 16-DEC-1993; 93WO-FR01260.
XX
PR 16-DEC-1992; 92FR-0015201.
XX
PA (LJOU) INST RECH JOUVEINAL.
XX
PI Benicourt C, Blanchard C, Junien J;
XX
DR WPI; 1994-217890/26.
DR N-PSDB; AAQ68388.
XX

```
PT   Recombinant canine gastric lipase and nucleic acid encoding it -
PT   are used for improving absorption of ingested fat, treating
PT   mucoviscidosis etc. and in enzymatic bio-conversions
XX
PS   Claim 13; Fig 9A; 52pp; French.
XX
CC   The sequence given below is the sequence of figure 9A, altered
CC   according to the amendments described on page 2 of the appended
CC   letter.
CC   CGL is used to improve absorption of ingested fat, in healthy and
CC   sick patients (e.g. having altered levels of gastric lipase); to
CC   treat conditions associated with insufficiency (or lack) of lipases,
CC   esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC   where immobilised, for bioconversions, e.g. hydrolysis or
CC   transesterification (other mammalian gastric lipases, or derivs.,
CC   can be used in this application).
XX
SQ   Sequence   379 AA;


  Query Match             50.8%;  Score 1141.5;  DB 15;  Length 379;
  Best Local Similarity   56.5%;  Pred. No. 1.4e-107;
  Matches  210;  Conservative   67;  Mismatches   90;  Indels    5;  Gaps    3;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| ||||| ||||||| ::||| |    :  | |||  |
Db       7 pt---npevtmnisqmitywgypaeeyevvtedgyilgidripygrknsenigrrpvafl 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |:|||||||||| |||||||:|||:|||||| |:|::   | |  ||||||:
Db      64 qhgllasatnwisnlpnnslafiladagydvwlgnsrgntwarrnlyyspdsvefwafsf 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:|||:||||:|::|||:|||||:|||||| |:|||::|| ::||||:|
Db     124 demakydlpatidfilkktgqdklhyvghsqgttigfiafstnpklakriktfyalapva 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLL 275
           |||: ::   | :|:|  : | :|| | | |   | | |   :| :  :| :||| : :
Db     184 tvkytetllnklmlvpsflfklifgnkif-yphhffdqflatevcsretvdllcsnalfi 242

Qy     276 LGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQP 335
           : ||:| |:|||  || :|  |||||||:|||||| ||: :|||||  :|:   :|
Db     243 icgfdtmnlnmsrldvylshnpagtsvqnvlhwsqavksgkfqafdwgspvqnmmhyhqs 302

Qy     336 TPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDA 395
             | | : || || |:| || | |::| || :|||:: |||||: || : |:|||| :||
Db     303 mppyynltdmhvpiavwnggndlladphdvdlllsklpnliyhrkippynhldfiwamda 362

Qy     396 PHRMYNEIIHLM 407
           |  :|||||: :|
Db     363 pqavyneivsmm 374


RESULT   9
AAW09382
ID   AAW09382 standard; Protein; 379 AA.
XX
AC   AAW09382;
XX
DT   19-AUG-1997  (first entry)
XX
DE   Dog gastric lipase protein sequence.
XX
KW   Dog; duodenal; gastric; lipase; transgenic; plant; recombinant; extract;
KW   food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;
KW   hydrolysis; trans-esterification; substrate; enzyme; biofuel.
XX
OS   Canis familiaris.
XX
PN   WO9633277-A2.
XX
PD   24-OCT-1996.
XX
PF   19-APR-1996;   96WO-FR00606.
XX
PR   20-APR-1995;   95FR-0004754.
XX
PA   (BIOC-) BIOCEM SA.
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Baudino S,  Benicourt C,  Cudrey C,  Gruber V,  Lenee P;
PI   Merot B;
XX
DR   WPI; 1996-485783/48.
DR   N-PSDB; AAT58915.
XX
PT   Recombinant human or canine pre-duodenal lipase prodn. in transgenic
PT   plants - useful for facilitating absorption of fat, as bio-catalysts
PT   and for prodn. of bio-fuel
XX
PS   Claim 2; Fig 2; 130pp; French.
XX
CC   This is the amino acid sequence of the dog pre-duodenal (i.e. gastric)
CC   lipase enzyme.  The sequence can be used to generate transgenic plants
CC   producing recombinant lipase in an enzymatically active form.
CC   Alternatively the first 4 or 54 amino acids of this sequence can be
CC   deleted to form the derivatives designated delta-4 or delta-54
CC   respectively.  Plants, or their extracts, expressing the lipases or the
CC   truncated derivatives, can be used: (a) as pharmaceuticals or food to
CC   facilitate absorption of fat, either in healthy subjects or in patients
CC   with inadequate levels of gastric/pancreatic lipase e.g. those with
CC   cystic fibrosis or exocrine pancreatic insufficient, in the elderly or in
CC   treatment which alters fat adsorption; (b) for performing industrial or
CC   agricultural reactions, e.g. in processing of fats or in the dairy
CC   industry, for hydrolysis or transesterification reactions, etc., where
CC   the plant material may provide both enzyme and substrate.  The transgenic
CC   plants can also be used for biofuel production.
XX
SQ   Sequence   379 AA;


  Query Match             50.8%;  Score 1141.5;  DB 17;  Length 379;
  Best Local Similarity   56.5%;  Pred. No. 1.4e-107;
  Matches  210;  Conservative   67;  Mismatches   90;  Indels    5;  Gaps    3;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| ||||| ||||||| ::||| |    :  | |||  |
Db       7 pt---npevtmnisqmitywgypaeeyevvtedgyilgidripygrknsenigrrpvafl 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |:|||||||||| |||||||:|||:|||||| |:|::   | |  ||||||:
Db      64 qhgllasatnwisnlpnnslafiladagydvwlgnsrgntwarrnlyyspdsvefwafsf 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:|||:||||:|::|||:|||||:|||||| |:|||::|| ::||||:|
Db     124 demakydlpatidfilkktgqdklhyvghsqgttigfiafstnpklakriktfyalapva 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLL 275
           |||: ::   | :|:|  : | :|| | | |   | | |   :| :  :| :||| : :
Db     184 tvkytetllnklmlvpsflfklifgnkif-yphhffdqflatevcsretvdllcsnalfi 242

Qy     276 LGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQP 335
           : ||:| |:|||  || :|  |||||||:|||||| ||: :|||||  :|:   :|
Db     243 icgfdtmnlnmsrldvylshnpagtsvqnvlhwsqavksgkfqafdwgspvqnmmhyhqs 302

Qy     336 TPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDA 395
             | | : || || |:| || | |::| || :|||:: |||||: || : |:|||| :||
Db     303 mppyynltdmhvpiavwnggndlladphdvdlllsklpnliyhrkippynhldfiwamda 362

Qy     396 PHRMYNEIIHLM 407
           |  :|||||: :|
Db     363 pqavyneivsmm 374


RESULT  10
AAR56871
```

ID AAR56871 standard; Protein; 380 AA.
XX
AC AAR56871;
XX
DT 20-FEB-1995 (first entry)
XX
DE Canine gastric lipase (Met-CGL).
XX
KW Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
KW bioconversion; exocrine pancreatic insufficiency.
XX
OS Canis familiaris.
XX
PN WO9413816-A.
XX
PD 23-JUN-1994.
XX
PF 16-DEC-1993; 93WO-FR01260.
XX
PR 16-DEC-1992; 92FR-0015201.
XX
PA (LJOU) INST RECH JOUVEINAL.
XX
PI Benicourt C, Blanchard C, Junien J;
XX
DR WPI; 1994-217890/26.
DR N-PSDB; AAQ68389.
XX
PT Recombinant canine gastric lipase and nucleic acid encoding it -
PT are used for improving absorption of ingested fat, treating
PT mucoviscidosis etc. and in enzymatic bio-conversions
XX
PS Claim 14; Fig 9A; 52pp; French.
XX
CC The sequence given below is the sequence of figure 9A, altered
CC according to the amendments described on page 2 of the appended
CC letter.
CC CGL is used to improve absorption of ingested fat, in healthy and
CC sick patients (e.g. having altered levels of gastric lipase); to
CC treat conditions associated with insufficiency (or lack) of lipases,
CC esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC where immobilised, for bioconversions, e.g. hydrolysis or
CC transesterification (other mammalian gastric lipases, or derivs.,
CC can be used in this application).
XX
SQ Sequence 380 AA;

Query Match 50.8%; Score 1141.5; DB 15; Length 380;
Best Local Similarity 56.5%; Pred. No. 1.4e-107;
Matches 210; Conservative 67; Mismatches 90; Indels 5; Gaps 3;

Qy 37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
Db 8 pt---npevtmnisqmitywgypaeeyevvtedgyilgidripygrknsenigrrpvafl 64

Qy 97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
Db 65 qhgllasatnwisnlpnnslafiladagydvwlgnsrgntwarrnlyyspdsvefwafsf 124

Qy 157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
Db 125 demakydlpatidfilkktgqdklhyvghsqgttigfiafstnpklakriktfyalapva 184

Qy 217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLL 275
Db 185 tvkytetllnklmlvpsflfklifgnkif-yphhffdqflatevcsretvdllcsnalfi 243

Qy 276 LGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQP 335
Db 244 icgfdtmnlnmsrldvylshnpagtsvqnvlhwsqavksgkfqafdwgspvqnmmhyhqs 303

Qy 336 TPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDA 395
Db 304 mppyynltdmhvpiavwnggndlladphdvdlllsklpnliyhrkippynhldfiwamda 363

Qy 396 PHRMYNEIIHLM 407
Db 364 pqavyneivsmm 375

RESULT 11
AAP60724
ID AAP60724 standard; Protein; 398 AA.
XX
AC AAP60724;
XX
DT 08-JUN-1991 (first entry)
XX
DE Sequence of pregastric lipase.
XX
KW Precursor polypeptide; secretion vector; enzyme.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT /label= signal
FT Protein 20..398
XX
PN WO8603778-A.
XX
PD 03-JUL-1986.
XX
PF 23-DEC-1985; 85WO-GB00599.
XX
PR 21-DEC-1984; 84GB-0032483.
PR 23-DEC-1985; 85WO-GB00599.
PR 01-JAN-1986; 86GB-0019568.
XX
PA (BREW-) BREWING RES FOUND.
PA (TUBB/) TUBB R S.
XX
PI Tubb RS;
XX
DR WPI; 1986-182910/28.
DR N-PSDB; AAN60685.
XX
PT New precursor polypeptide of defined sequence - and corresp. DNA
PT used to transform hosts for prodn. of the polypeptide
XX
PS Example; Fig 5; 60pp; English.
XX
CC The inventors claim a DNA sequence encoding AAP60723 linked to a
CC promoter upstream and a gene for a polypeptide downstream.
CC Particular examples are the yeast emzyme AMG, the mammalian enzyme,
CC gastric lipase and the mammalian lymphokine, interferon-alpha2.
XX
SQ Sequence 398 AA;

Query Match 49.8%; Score 1119; DB 7; Length 398;
Best Local Similarity 52.2%; Pred. No. 3e-105;
Matches 210; Conservative 73; Mismatches 107; Indels 12; Gaps 2;

Qy 17 MWLLILVAYMFQRNVNSVHMPTKAV-------DPEAFMNISEIIQHQGYPCEEYEVATED 69
Db 1 mwllltmasli-----svlgtthglfgklhpgspevtmnisqmitywgypneeyevvted 55

Qy 70 GYILSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWM 129
Db 56 gyilevnripygkknsgntgqrpvvflqhgllasatnwisnlpnnslafiladagydvwl 115

Qy 130 GNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGT 189
Db 116 gnsrgntwarrnlyyspdsvefwafsfdemakydlpatidfivkktgqkqlhyvghsqgt 175

```
Qy    190 TMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQT 249
          |:|||||||  |  ||::|| ::||||:||||: ||    |   :|  : | :|| | |
Db    176 tigfiafstnpslakriktfyalapvatvkytkslinklrfvpqslfkfifgdkifyphn 235

Qy    250 RFLRQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWS 309
          | : |   :| : :|: :||| : :: ||:: | | ||  || :|  |||||||: ||:
Db    236 ffdqflatevcsremlnllcsnalfiicgfdsknfntsrldvylshnpagtsvqnmfhwt 295

Qy    310 QAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLL 369
          ||| ||: :|:|||| :|    :|  |  | |  | || |:| ||:| |::|:|| :||
Db    296 qavksgkfqaydwgspvqnrmhydqsqppyynvtamnvpiavwnggkdlladpqdvglll 355

Qy    370 SEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQEE 411
          .:: |||||| || : |:|||| :|||  :||:|: :: :::
Db    356 pklpnliyhkeipfynhldfiwamdapqevyndivsmisedk 397


RESULT  12
AAP60658
ID   AAP60658 standard; Protein; 398 AA.
XX
AC   AAP60658;
XX
DT   22-AUG-1991  (first entry)
XX
DE   Sequence of human pregastric lipase.
XX
KW   Cystic fibrosis therapy; enzyme; lipase deficiency.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT   Protein         20..398
XX
PN   WO8601532-A.
XX
PD   13-MAR-1986.
XX
PF   15-AUG-1985;   85WO-GB00364.
XX
PR   21-AUG-1984;   84GB-0021210.
PR   15-AUG-1985;   85WO-GB00364.
PR   01-JAN-1986;   86GB-0008897.
XX
PA   (CELL-) CELLTECH LTD.
PA   (LOWE/) LOWE P A.
XX
PI   Lowe PA;
XX
DR   WPI; 1986-081634/12.
DR   N-PSDB; AAN60566.
XX
PT   New gastric lipase protein, esp. of human origin - for treating
PT   lipase deficiency, and DNA sequences coding for it
XX
PS   Disclosure; Fig 3; 39pp; English.
XX
CC   The inventors claim a pregastric lipase protein and a gene encoding
CC   it. Gastric lipase protein is useful for oral administration to
CC   treat lipase deficiency, e.g. cystic fibrosis or pancreatitis.
XX
SQ   Sequence   398 AA;


  Query Match             49.8%;  Score 1119;  DB 7;  Length 398;
  Best Local Similarity   52.2%;  Pred. No. 3e-105;
  Matches  210;  Conservative   73;  Mismatches  107;  Indels   12;  Gaps    2;

Qy     17 MWLLILVAYMFQRNVNSVHMPTKAV-------DPEAFMNISEIIQHQGYPCEEYEVATED 69
          ||||: :| :      ||   |  :          ||  ||||::| : ||| ||||| |||
Db      1 mwllltmasli-----svlgtthglfgklhpgspevtmnisqmitywgypneeyevvted 55

Qy     70 GYILSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWM 129
          |||| ||||| |      || |||| |||||:  |:||||||||||| |||||||:|||:
Db     56 gyilevnripygkknsgntgqrpvvflqhgllasatnwisnlpnnslafiladagydvwl 115

Qy    130 GNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGT 189
          |||||| |:|::   | |  ||||||:||||::|||| |:||::||||::::|||:||||
Db    116 gnsrgntwarrnlyyspdsvefwafsfdemakydlpatidfivkktgqkqlhyvghsqgt 175

Qy    190 TMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQT 249
          |:|||||||  | ||::|| ::||||:||||: ||    |   :|  : | :|| | |
Db    176 tigfiafstnpslakriktfyalapvatvkytkslinklrfvpqslfkfifgdkifyphn 235

Qy    250 RFLRQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWS 309
          | : |   :| : :|: :||| : :: ||:: | | ||  || :|  |||||||: ||:
Db    236 ffdqflatevcsremlnllcsnalfiicgfdsknfntsrldvylshnpagtsvqnmfhwt 295

Qy    310 QAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLL 369
          ||| ||: :|:|||| :|    :|  |  | |  | || |:| ||:| |::|:|| :||
Db    296 qavksgkfqaydwgspvqnrmhydqsqppyynvtamnvpiavwnggkdlladpqdvglll 355

Qy    370 SEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQEE 411
          :: |||||| || : |:|||| :|||  :||:|: :: :::
Db    356 pklpnliyhkeipfynhldfiwamdapqevyndivsmisedk 397


RESULT  13
AAW09383
ID   AAW09383 standard; Protein; 398 AA.
XX
AC   AAW09383;
XX
DT   19-AUG-1997  (first entry)
XX
DE   Human gastric lipase protein sequence.
XX
KW   Duodenal; gastric; lipase; transgenic; plant; recombinant; extract;
KW   food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;
KW   hydrolysis; trans-esterification; substrate; enzyme; biofuel; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "signal peptide"
FT   Protein         20..999
FT                   /note= "mature protein"
XX
PN   WO9633277-A2.
XX
PD   24-OCT-1996.
XX
PF   19-APR-1996;   96WO-FR00606.
XX
PR   20-APR-1995;   95FR-0004754.
XX
PA   (BIOC-) BIOCEM SA.
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Baudino S,  Benicourt C,  Cudrey C,  Gruber V,  Lenee P;
PI   Merot B;
XX
DR   WPI; 1996-485783/48.
DR   N-PSDB; AAT58916.
XX
PT   Recombinant human or canine pre-duodenal lipase prodn. in transgenic
PT   plants - useful for facilitating absorption of fat, as bio-catalysts
PT   and for prodn. of bio-fuel
XX
PS   Claim 3; Fig 5; 130pp; French.
XX
CC   This is the amino acid sequence of the human pre-duodenal (i.e. gastric)
CC   lipase enzyme.  The sequence can be used to generate transgenic plants
```

```
CC  producing recombinant lipase in an enzymatically active form.
CC  Alternatively amino acids 20-23 or 20-73 (1-4 or 1-54 of the mature
CC  protein, respectively) can be deleted to form the derivatives designated
CC  delta-4 or delta-54 respectively.  Plants, or their extracts, expressing
CC  the lipases or the truncated derivatives, can be used:
CC  (a) as pharmaceuticals or food to facilitate absorption of fat, either
CC  in healthy subjects or in patients with inadequate levels of
CC  gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine
CC  pancreatic insufficient, in the elderly or in patients undergoing medical
CC  treatment which alters fat adsorption; (b) for performing industrial or
CC  agricultural reactions, e.g. in processing of fats or in the dairy
CC  industry, for hydrolysis or transesterification reactions, etc., where
CC  the plant material may provide both enzyme and substrate.  The transgenic
CC  plants can also be used for biofuel production.
XX
SQ  Sequence  398 AA;

  Query Match             49.8%;  Score 1119;  DB 17;  Length 398;
  Best Local Similarity   52.2%;  Pred. No. 3e-105;
  Matches  210;  Conservative   73;  Mismatches  107;  Indels   12;  Gaps    2;

Qy     17 MWLLILVAYMFQRNVNSVHMPTKAV-------DPEAFMNISEIIQHQGYPCEEYEVATED 69
Db      1 mwllltmasli-----svlgtthglfgklhpgspevtmnisqmitywgypneeyevvted 55

Qy     70 GYILSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWM 129
Db     56 gyilevnripygkknsgntgqrpvvflqhgllasatnwisnlpnnslafiladagydvwl 115

Qy    130 GNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGT 189
Db    116 gnsrgntwarrnlyyspdsvefwafsfdemakydlpatidfivkktgqkqlhyvghsqgt 175

Qy    190 TMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQT 249
Db    176 tigfiafstnpslakriktfyalapvatvkytkslinklrfvpqslfkfifgdkifyphn 235

Qy    250 RFLRQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWS 309
Db    236 ffdqflatevcsremlnllcsnalfiicgfdsknfntsrldvylshnpagtsvqnmfhwt 295

Qy    310 QAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLL 369
Db    296 qavksgkfqaydwgspvqnrmhydqsqppyynvtamnvpiavwnggkdlladpqdvglll 355

Qy    370 SEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQEE 411
Db    356 pklpnliyhkeipfynhldfiwamdapqevyndivsmisedk 397

RESULT  14
AAB66086
ID   AAB66086 standard; Protein; 398 AA.
XX
AC   AAB66086;
XX
DT   30-MAR-2001  (first entry)
XX
DE   Human lipase protein.
XX
KW   Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
KW   central nervous system; focal brain disorder; bipolar affective disorder;
KW   global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW   senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW   Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW   neuropsychiatric; psychoactive substance use; anxiety.
XX
OS   Homo sapiens.
XX
PN   WO200077239-A2.
XX
PD   21-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14858.
XX
PR   14-JUN-1999;   99US-0333159.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   McCarthy SA,  Fraser CC,  Sharp JD,  Barnes TM;
XX
DR   WPI; 2001-032313/04.
XX
PT   TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT   screening assays and diagnostic assays and for the treatment of
PT   neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT   disease -
XX
PS   Disclosure; Fig 6; 359pp; English.
XX
CC   The present invention relates to TANGO or INTERCEPT proteins and coding
CC   sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC   AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC   sequences are useful for the treatment of neurological disorders such as
CC   central nervous system (CNS) disorders, CNS-related disorders, focal
CC   brain disorders, global-diffuse cerebral disorders and other
CC   neurological and cerebrovascular disorders. The CNS disorders include
CC   Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC   lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC   autonomic function disorders such as hypertension and sleep disorders,
CC   neuropsychiatric disorders, psychoactive substance use disorders,
CC   anxiety, and bipolar affective disorder. The present sequence is a
CC   sequence used in a sequence homology alignment with the TANGO/INTERCEPT
CC   sequences of the present invention.
XX
SQ   Sequence  398 AA;

  Query Match             49.8%;  Score 1119;  DB 22;  Length 398;
  Best Local Similarity   52.2%;  Pred. No. 3e-105;
  Matches  210;  Conservative   73;  Mismatches  107;  Indels   12;  Gaps    2;

Qy     17 MWLLILVAYMFQRNVNSVHMPTKAV-------DPEAFMNISEIIQHQGYPCEEYEVATED 69
Db      1 mwllltmasli-----svlgtthglfgklhpgspevtmnisqmitywgypneeyevvted 55

Qy     70 GYILSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWM 129
Db     56 gyilevnripygkknsgntgqrpvvflqhgllasatnwisnlpnnslafiladagydvwl 115

Qy    130 GNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGT 189
Db    116 gnsrgntwarrnlyyspdsvefwafsfdemakydlpatidfivkktgqkqlhyvghsqgt 175

Qy    190 TMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQT 249
Db    176 tigfiafstnpslakriktfyalapvatvkytkslinklrfvpqslfkfifgdkifyphn 235

Qy    250 RFLRQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWS 309
Db    236 ffdqflatevcsremlnllcsnalfiicgfdsknfntsrldvylshnpagtsvqnmfhwt 295

Qy    310 QAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLL 369
Db    296 qavksgkfqaydwgspvqnrmhydqsqppyynvtamnvpiavwnggkdlladpqdvglll 355

Qy    370 SEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQEE 411
Db    356 pklpnliyhkeipfynhldfiwamdapqevyndivsmisedk 397

RESULT  15
AAB66070
ID   AAB66070 standard; Protein; 144 AA.
XX
```

AC AAB66070;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294 cytoplasmic domain.
XX
KW TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
XX
PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease -
XX
PS Claim 8; Page 327; 359pp; English.
XX
CC The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other
CC neurological and cerebrovascular disorders. The CNS disorders include
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC autonomic function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder.
XX
SQ Sequence 144 AA;

Query Match 34.8%; Score 783; DB 22; Length 144;
Best Local Similarity 100.0%; Pred. No. 1e-71;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    280 NTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVR 339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ntnnmnmsrasvyaahtlagtsvqnilhwsqavnsgelrafdwgsetknlekcnqptpvr 60

Qy    340 YRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHRM 399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 yrvrdmtvptamwtggqdwlsnpedvkmllsevtnliyhknipewahvdfiwgldaphrm 120

Qy    400 YNEIIHLMQQEETNLSQGRCEAVL 423
          ||||||||||||||||||||||||
Db    121 yneiihlmqqeetnlsqgrceavl 144
```

Search completed: January 30, 2002, 23:00:02
Job time: 3043 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 30, 2002, 22:11:19 ; Search time 65.9 Seconds
(without alignments)
144.444 Million cell updates/sec

Title: US-09-578-063-47
Perfect score: 2247
Sequence: 1 MLETLSRQWIVSHRMEMWLL.....IHLMQQEETNLSQGRCEAVL 423

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1161	51.7	377	1	US-08-227-108-17	Sequence 17, Appl
2	1161	51.7	377	2	US-09-073-674-17	Sequence 17, Appl
3	1150.5	51.2	379	1	US-08-227-108-18	Sequence 18, Appl
4	1150.5	51.2	379	2	US-09-073-674-18	Sequence 18, Appl
5	1141.5	50.8	379	1	US-08-227-108-3	Sequence 3, Appli
6	1141.5	50.8	379	2	US-09-073-674-3	Sequence 3, Appli
7	1141.5	50.8	380	1	US-08-227-108-5	Sequence 5, Appli
8	1141.5	50.8	380	2	US-09-073-674-5	Sequence 5, Appli
9	1098.5	48.9	380	1	US-08-227-108-16	Sequence 16, Appl
10	1098.5	48.9	380	2	US-09-073-674-16	Sequence 16, Appl
11	105	4.7	430	1	US-08-484-105-22	Sequence 22, Appl
12	105	4.7	430	1	US-08-484-106-22	Sequence 22, Appl
13	102	4.5	346	2	US-08-602-359A-34	Sequence 34, Appl
14	98.5	4.4	388	1	US-08-232-519-2	Sequence 2, Appli
15	98.5	4.4	388	1	US-08-456-956-2	Sequence 2, Appli
16	89.5	4.0	624	2	US-08-756-317-9	Sequence 9, Appli
17	89	4.0	1674	2	US-08-968-542C-12	Sequence 12, Appl
18	88.5	3.9	869	2	US-08-483-101-15	Sequence 15, Appl
19	84.5	3.8	338	2	US-08-602-359A-40	Sequence 40, Appl
20	84.5	3.8	935	1	US-07-707-367-2	Sequence 2, Appli
21	84	3.7	719	4	US-09-386-607-2	Sequence 2, Appli
22	82.5	3.7	485	2	US-08-446-803-1	Sequence 1, Appli
23	82.5	3.7	485	2	US-08-861-837-1	Sequence 1, Appli
24	82.5	3.7	485	2	US-08-600-908A-12	Sequence 12, Appl
25	82.5	3.7	485	3	US-08-683-838A-12	Sequence 12, Appl
26	82.5	3.7	485	3	US-08-600-656-1	Sequence 1, Appli
27	82.5	3.7	485	4	US-09-170-670-1	Sequence 1, Appli
28	82.5	3.7	485	4	US-09-170-670-7	Sequence 7, Appli
29	82.5	3.7	485	4	US-09-193-068-1	Sequence 1, Appli
30	82.5	3.7	485	4	US-09-193-068-7	Sequence 7, Appli
31	82.5	3.7	485	4	US-09-183-412-1	Sequence 1, Appli
32	82.5	3.7	485	4	US-09-183-412-7	Sequence 7, Appli
33	82.5	3.7	485	4	US-09-264-097-7	Sequence 7, Appli
34	82.5	3.7	485	4	US-09-354-191A-1	Sequence 1, Appli
35	82	3.6	330	2	US-08-997-080-30	Sequence 30, Appl
36	82	3.6	330	2	US-08-997-362-30	Sequence 30, Appl
37	82	3.6	330	3	US-08-873-970-30	Sequence 30, Appl
38	82	3.6	330	4	US-09-095-855-30	Sequence 30, Appl
39	82	3.6	330	4	US-08-705-347A-30	Sequence 30, Appl
40	82	3.6	493	1	US-07-615-448A-7	Sequence 7, Appli
41	82	3.6	493	1	US-08-196-361-7	Sequence 7, Appli
42	82	3.6	493	2	US-08-446-934-7	Sequence 7, Appli
43	82	3.6	493	2	US-08-448-128-7	Sequence 7, Appli
44	82	3.6	493	3	US-08-948-703-7	Sequence 7, Appli
45	81.5	3.6	428	1	US-08-190-802A-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-08-227-108-17
; Sequence 17, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 377 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-227-108-17

Query Match 51.7%; Score 1161; DB 1; Length 377;
Best Local Similarity 57.7%; Pred. No. 2e-114;
Matches 213; Conservative 63; Mismatches 93; Indels 0; Gaps

Qy 42 DPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLLQHGLV 1
:||| |||::| : ||||:|||| ||||||| | ||| | : | |||| |||||:

```
Db      9 NPEANMNISQMITYWGYPCQEYEVVTEDGYILGVYRIPHGKNNSENIGKRPVVYLQHGLI 68

Qy    102 GGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEMAR 161
          |:|||:|||||| |:|||||:|||:|||||| |||:   | |  ||||||:||||:
Db     69 ASATNWIANLPNNSLAFMLADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAK 128

Qy    162 FDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVKHA 221
          :|||| ||||:||||||||:|||:|||||:|||||| | ||:||| ::||||:||||:
Db    129 YDLPATINFIVQKTGQEKIHYVGHSQGTTIGFIAFSTNPTLAKKIKTFYALAPVATVKYT 188

Qy    222 KSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGFNT 281
          :||  |   :|  : | :|||| ||  | |   |   :| : :|| :||| : :  ||:
Db    189 QSPLKKISFIPTFLFKLMFGKKMFLPHTYFDDFLGTEVCSREVLDLLCSNTLFIFCGFDK 248

Qy    282 NNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVRYR 341
           |:|:||  ||  |  ||||||: |||:| | ||: :||:||| ::|:   || ||  |
Db    249 KNLNVSRFDVYLGHNPAGTSVQDFLHWAQLVRSGKFQAFNWGSPSQNMLHYNQKTPPEYD 308

Qy    342 VRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYN 401
          |  |||| |:| || | |::|:|| ||| :::||::|| |  : |:|||| :|||  :||
Db    309 VSAMTVPVAVWNGGNDILADPQDVAMLLPKLSNLLFHKEILAYNHLDFIWAMDAPQEVYN 368

Qy    402 EIIHLMQQE 410
          |:| :| ::
Db    369 EMISMMAED 377


RESULT   2
US-09-073-674-17
; Sequence 17, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  377 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-073-674-17


  Query Match           51.7%;  Score 1161;  DB 2;  Length 377;
  Best Local Similarity 57.7%;  Pred. No. 2e-114;
  Matches  213;  Conservative   63;  Mismatches   93;  Indels    0;  Gaps    0;

Qy     42 DPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLLQHGLV 101
          :||| ||||::| : ||||:|||| ||||||| | ||| |    :  | |||| |||||:
Db      9 NPEANMNISQMITYWGYPCQEYEVVTEDGYILGVYRIPHGKNNSENIGKRPVVYLQHGLI 68

Qy    102 GGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEMAR 161
          |:|||:|||||| |:|||||:|||:|||||| ||||:   | |  ||||||:||||:
Db     69 ASATNWIANLPNNSLAFMLADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAK 128

Qy    162 FDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVKHA 221
          :|||| ||||:||||||||:|||:|||||:|||||| | ||:||| ::||||:||||:
Db    129 YDLPATINFIVQKTGQEKIHYVGHSQGTTIGFIAFSTNPTLAKKIKTFYALAPVATVKYT 188

Qy    222 KSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGFNT 281
          :||  |   :|  : | :|||| ||  | |   |   :| : :|| :||| : :  ||:
Db    189 QSPLKKISFIPTFLFKLMFGKKMFLPHTYFDDFLGTEVCSREVLDLLCSNTLFIFCGFDK 248

Qy    282 NNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVRYR 341
           |:|:||  ||  |  ||||||: |||:| | ||: :||:||| ::|:   || ||  |
Db    249 KNLNVSRFDVYLGHNPAGTSVQDFLHWAQLVRSGKFQAFNWGSPSQNMLHYNQKTPPEYD 308

Qy    342 VRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYN 401
          |  |||| |:| || | |::|:|| ||| :::||::|| |  : |:|||| :|||  :||
Db    309 VSAMTVPVAVWNGGNDILADPQDVAMLLPKLSNLLFHKEILAYNHLDFIWAMDAPQEVYN 368

Qy    402 EIIHLMQQE 410
          |:| :| ::
Db    369 EMISMMAED 377


RESULT   3
US-08-227-108-18
; Sequence 18, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  379 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-227-108-18
```

```
  Query Match             51.2%;  Score 1150.5;  DB 1;  Length 379;
  Best Local Similarity   56.8%;  Pred. No. 2.6e-113;
  Matches  213;  Conservative   65;  Mismatches   94;  Indels    3;  Gaps    1;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| |:||| ||||||| ||||| |       | |||| |
Db       7 PT---NPEVNMNISQMISYWGYPSEKYEVVTEDGYILEVNRIPYGKKNSGNRGQRPVVFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |||||||||||||| |||||||: ||:|||||| |||::   | |  ||||||:
Db      64 QHGLLASASNWISNLPNNSLAFILADAGYGVWLGNSRGNTWSRRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:||:::|||||::|||:|||||:||||||| |:||::|| ::|||||:|
Db     124 DEMAKYDLPATIDFIVKETGQEKLHYVGHSQGTTIGFIAFSTNPKLAERIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLL 276
           |||: ||   |  :|  | | :|| | |     | : |   :| :  |: |||| : ::
Db     184 TVKYTKSLVNKLRFIPPTMFKIIFGDKIFYPHNFFDQFLATQVCSRETLNVICSNALFII 243

Qy     277 GGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPT 336
           ||:: |:|||  || :|  |||||||:|||:||| ||  :||:|||  :|:   ||||
Db     244 CGFDSANLNMSRLDVYVSHNPAGTSVQNMLHWTQAVKSGNFQAFNWGSPAQNVVHFNQPT 303

Qy     337 PVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAP 396
           |  | |  | || |:|:|| |||::|:|| :|| :::|||||| |  : |:|||| ::||
Db     304 PPYYNVTAMNVPIAVWSGGNDWLADPQDVDLLLPKLSNLIYHKEILPYNHLDFIWAMNAP 363

Qy     397 HRMYNEIIHLMQQEE 411
             :||||| :| :::
Db     364 QEVYNEIISMMAKDK 378


RESULT   4
US-09-073-674-18
; Sequence 18, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  379 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-073-674-18


  Query Match             51.2%;  Score 1150.5;  DB 2;  Length 379;
  Best Local Similarity   56.8%;  Pred. No. 2.6e-113;
  Matches  213;  Conservative   65;  Mismatches   94;  Indels    3;  Gaps    1;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| |:||| ||||||| ||||| |       | |||| |
Db       7 PT---NPEVNMNISQMISYWGYPSEKYEVVTEDGYILEVNRIPYGKKNSGNRGQRPVVFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |||||||||||||| |||||||: ||:|||||| |||::   | |  ||||||:
Db      64 QHGLLASASNWISNLPNNSLAFILADAGYGVWLGNSRGNTWSRRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:||:::|||||::|||:|||||:||||||| |:||::|| ::|||||:|
Db     124 DEMAKYDLPATIDFIVKETGQEKLHYVGHSQGTTIGFIAFSTNPKLAERIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLL 276
           |||: ||   |  :|  | | :|| | |     | : |   :| :  |: |||| : ::
Db     184 TVKYTKSLVNKLRFIPPTMFKIIFGDKIFYPHNFFDQFLATQVCSRETLNVICSNALFII 243

Qy     277 GGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPT 336
           ||:: |:|||  || :|  |||||||:|||:||| ||  :||:|||  :|:   ||||
Db     244 CGFDSANLNMSRLDVYVSHNPAGTSVQNMLHWTQAVKSGNFQAFNWGSPAQNVVHFNQPT 303

Qy     337 PVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAP 396
           |  | |  | || |:|:|| |||::|:|| :|| :::|||||| |  : |:|||| ::||
Db     304 PPYYNVTAMNVPIAVWSGGNDWLADPQDVDLLLPKLSNLIYHKEILPYNHLDFIWAMNAP 363

Qy     397 HRMYNEIIHLMQQEE 411
             :||||| :| :::
Db     364 QEVYNEIISMMAKDK 378


RESULT   5
US-08-227-108-3
; Sequence 3, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
```

```
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  379 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-227-108-3


  Query Match             50.8%;  Score 1141.5;  DB 1;  Length 379;
  Best Local Similarity   56.5%;  Pred. No. 2.4e-112;
  Matches  210;  Conservative   67;  Mismatches   90;  Indels    5;  Gaps    3;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| ||||| ||||||| ::||| |    :  | |||  |
Db       7 PT---NPEVTMNISQMITYWGYPAEEYEVVTEDGYILGIDRIPYGRKNSENIGRRPVAFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |:||||||||||| |||||||:|||:|||||| |:|::   | |  ||||||:
Db      64 QHGLLASATNWISNLPNNSLAFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:|||:||||:|::|||:|||||:||||||| |:||::|| ::||||:|
Db     124 DEMAKYDLPATIDFILKKTGQDKLHYVGHSQGTTIGFIAFSTNPKLAKRIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLL 275
           |||: ::   | :|:|  : | :|| | | |   | | |   :| :  :| :||| : :
Db     184 TVKYTETLLNKLMLVPSFLFKLIFGNKIF-YPHHFFDQFLATEVCSRETVDLLCSNALFI 242

Qy     276 LGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQP 335
           : ||:| |:|||| || :|  ||||||:|||||| ||: :|||||  :|:   :|
Db     243 ICGFDTMNLNMSRLDVYLSHNPAGTSVQNVLHWSQAVKSGKFQAFDWGSPVQNMMHYHQS 302

Qy     336 TPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDA 395
            |  | : || || |:| || | |::| || :|||:: ||||: || : |:|||| :||
Db     303 MPPYYNLTDMHVPIAVWNGGNDLLADPHDVDLLLSKLPNLIYHRKIPPYNHLDFIWAMDA 362

Qy     396 PHRMYNEIIHLM 407
           |  :||||: :|
Db     363 PQAVYNEIVSMM 374


RESULT   6
US-09-073-674-3
; Sequence 3, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  379 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-073-674-3


  Query Match             50.8%;  Score 1141.5;  DB 2;  Length 379;
  Best Local Similarity   56.5%;  Pred. No. 2.4e-112;
  Matches  210;  Conservative   67;  Mismatches   90;  Indels    5;  Gaps    3;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| ||||| ||||||| ::||| |    :  | |||  |
Db       7 PT---NPEVTMNISQMITYWGYPAEEYEVVTEDGYILGIDRIPYGRKNSENIGRRPVAFL 63

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |:||||||||||| |||||||:|||:|||||| |:|::   | |  ||||||:
Db      64 QHGLLASATNWISNLPNNSLAFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSF 123

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:|||:||||:|::|||:|||||:||||||| |:||::|| ::||||:|
Db     124 DEMAKYDLPATIDFILKKTGQDKLHYVGHSQGTTIGFIAFSTNPKLAKRIKTFYALAPVA 183

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLL 275
           |||: ::   | :|:|  : | :|| | | |   | | |   :| :  :| :||| : :
Db     184 TVKYTETLLNKLMLVPSFLFKLIFGNKIF-YPHHFFDQFLATEVCSRETVDLLCSNALFI 242

Qy     276 LGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQP 335
           : ||:| |:|||| || :|  ||||||:|||||| ||: :|||||  :|:   :|
Db     243 ICGFDTMNLNMSRLDVYLSHNPAGTSVQNVLHWSQAVKSGKFQAFDWGSPVQNMMHYHQS 302

Qy     336 TPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDA 395
            |  | : || || |:| || | |::| ||.:|||:: ||||: || : |:|||| :||
Db     303 MPPYYNLTDMHVPIAVWNGGNDLLADPHDVDLLLSKLPNLIYHRKIPPYNHLDFIWAMDA 362

Qy     396 PHRMYNEIIHLM 407
           |  :||||: :|
Db     363 PQAVYNEIVSMM 374


RESULT   7
US-08-227-108-5
; Sequence 5, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
```

```
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  380 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-227-108-5


  Query Match           50.8%;  Score 1141.5;  DB 1;  Length 380;
  Best Local Similarity   56.5%;  Pred. No. 2.4e-112;
  Matches  210;  Conservative   67;  Mismatches   90;  Indels    5;  Gaps    3;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| ||||| ||||||| ::||| |    :  | |||  |
Db       8 PT---NPEVTMNISQMITYWGYPAEEYEVVTEDGYILGIDRIPYGRKNSENIGRRPVAFL 64

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |:||||||||||| |||||||:|||:|||||| |:|::   | |  ||||||:
Db      65 QHGLLASATNWISNLPNNSLAFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSF 124

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:|||:||||:|::|||:|||||:|||||| |:||::|| ::|||||:|
Db     125 DEMAKYDLPATIDFILKKTGQDKLHYVGHSQGTTIGFIAFSTNPKLAKRIKTFYALAPVA 184

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLL 275
           |||: ::   | :|:|  : | :|| | | |   | | |   :| :  :| :|||  : :
Db     185 TVKYTETLLNKLMLVPSFLFKLIFGNKIF-YPHHFFDQFLATEVCSRETVDLLCSNALFI 243

Qy     276 LGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQP 335
           : ||:| |:|||| || :|  |||||||:||||||| ||: :|||||  :|:   :|
Db     244 ICGFDTMNLNMSRLDVYLSHNPAGTSVQNVLHWSQAVKSGKFQAFDWGSPVQNMMHYHQS 303

Qy     336 TPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDA 395
             | | : || || |:| || | |::| || :|||:: |||||: || : |:|||| :||
Db     304 MPPYYNLTDMHVPIAVWNGGNDLLADPHDVDLLLSKLPNLIYHRKIPPYNHLDFIWAMDA 363

Qy     396 PHRMYNEIIHLM 407
           |  :||||: :|
Db     364 PQAVYNEIVSMM 375


RESULT   8
US-09-073-674-5
; Sequence 5, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  380 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-073-674-5


  Query Match           50.8%;  Score 1141.5;  DB 2;  Length 380;
  Best Local Similarity   56.5%;  Pred. No. 2.4e-112;
  Matches  210;  Conservative   67;  Mismatches   90;  Indels    5;  Gaps    3;

Qy      37 PTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLL 96
           ||   :||  ||||::| : ||| ||||| ||||||| ::||| |    :  | |||  |
Db       8 PT---NPEVTMNISQMITYWGYPAEEYEVVTEDGYILGIDRIPYGRKNSENIGRRPVAFL 64

Qy      97 QHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSY 156
           ||||:  |:||||||||||| |||||||:|||:|||||| |:|::   | |  ||||||:
Db      65 QHGLLASATNWISNLPNNSLAFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSF 124

Qy     157 DEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIA 216
           ||||::|||| |:|||:||||:|::|||:|||||:|||||| |:||::|| ::|||||:|
Db     125 DEMAKYDLPATIDFILKKTGQDKLHYVGHSQGTTIGFIAFSTNPKLAKRIKTFYALAPVA 184

Qy     217 TVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLL 275
           |||: ::   | :|:|  : | :|| | | |   | | |   :| :  :| :|||  : :
Db     185 TVKYTETLLNKLMLVPSFLFKLIFGNKIF-YPHHFFDQFLATEVCSRETVDLLCSNALFI 243

Qy     276 LGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQP 335
           : ||:| |:|||| || :|  |||||||:||||||| ||: :|||||  :|:   :|
Db     244 ICGFDTMNLNMSRLDVYLSHNPAGTSVQNVLHWSQAVKSGKFQAFDWGSPVQNMMHYHQS 303

Qy     336 TPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDA 395
             | | : || || |:| || | |::| || :|||:: |||||: || : |:|||| :||
Db     304 MPPYYNLTDMHVPIAVWNGGNDLLADPHDVDLLLSKLPNLIYHRKIPPYNHLDFIWAMDA 363

Qy     396 PHRMYNEIIHLM 407
           |  :||||: :|
Db     364 PQAVYNEIVSMM 375


RESULT   9
US-08-227-108-16
; Sequence 16, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  380 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-227-108-16


  Query Match             48.9%;  Score 1098.5;  DB 1;  Length 380;
  Best Local Similarity   54.6%;  Pred. No. 8.4e-108;
  Matches  202;  Conservative   69;  Mismatches   98;  Indels    1;  Gaps    1;

Qy      43 PEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLLQHGLVG 102
           ||  ||||::| : ||| ||||| ||||||| ||||| |      || |||| |||||:
Db      10 PEVTMNISQMITYWGYPNEEYEVVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLA 69

Qy     103 GASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFW-AFSYDEMAR 161
            |:||||||||||| |||||||:|||:|||||| |:|::   | |  ||| |||:||||:
Db      70 SATNWISNLPNNSLAFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAAFSFDEMAK 129

Qy     162 FDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVKHA 221
           :|||| |:||::|||||:::::|||:|||||:|||||| | ||::|| ::||||:||||:
Db     130 YDLPATIDFIVKKTGQKQLHYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYT 189

Qy     222 KSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGFNT 281
           ||   |   :|  : | :|| | |     | : |   :| : :|: :||| : :: ||::
Db     190 KSLINKLRFVPQSLFKFIFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDS 249

Qy     282 NNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVRYR 341
            | | ||  || :|  |||||||: ||:||| ||: :|:|||| :|   :|  |  |
Db     250 KNFNTSRLDVYLSHNPAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYN 309

Qy     342 VRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYN 401
           |  | || |:| ||:| |::|:|| :|| :: |||||| || : |:|||| :|||  :||
Db     310 VTAMNVPIAVWNGGKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYN 369

Qy     402 EIIHLMQQEE 411
           :|: :: :::
Db     370 DIVSMISEDK 379


RESULT  10
US-09-073-674-16
; Sequence 16, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  380 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-073-674-16


  Query Match             48.9%;  Score 1098.5;  DB 2;  Length 380;
  Best Local Similarity   54.6%;  Pred. No. 8.4e-108;
  Matches  202;  Conservative   69;  Mismatches   98;  Indels    1;  Gaps    1;

Qy      43 PEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLLQHGLVG 102
           ||  ||||::| : ||| ||||| ||||||| ||||| |      || |||| |||||:
Db      10 PEVTMNISQMITYWGYPNEEYEVVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLA 69

Qy     103 GASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFW-AFSYDEMAR 161
            |:||||||||||| |||||||:|||:|||||| |:|::   | |  ||| |||:||||:
Db      70 SATNWISNLPNNSLAFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAAFSFDEMAK 129

Qy     162 FDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVKHA 221
           :|||| |:||::|||||:::::|||:|||||:|||||| | ||::|| ::||||:||||:
Db     130 YDLPATIDFIVKKTGQKQLHYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYT 189

Qy     222 KSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGFNT 281
           ||   |   :|  : | :|| | |     | : |   :| : :|: :||| : :: ||::
Db     190 KSLINKLRFVPQSLFKFIFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDS 249

Qy     282 NNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVRYR 341
            | | ||  || :|  |||||||: ||:||| ||: :|:|||| :|   :|  |  |
Db     250 KNFNTSRLDVYLSHNPAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYN 309

Qy     342 VRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYN 401
           |  | || |:| ||:| |::|:|| :|| :: |||||| || : |:|||| :|||  :||
Db     310 VTAMNVPIAVWNGGKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYN 369

Qy     402 EIIHLMQQEE 411
           :|: :: :::
Db     370 DIVSMISEDK 379


RESULT  11
US-08-484-105-22
; Sequence 22, Application US/08484105
; Patent No. 5589341
;  GENERAL INFORMATION:
;    APPLICANT:  STILLMAN, Bruce
;    APPLICANT:  BELL, Stephen P
;    APPLICANT:  KOBAYASHI, Ryuji
;    APPLICANT:  RINE, Jasper
;    APPLICANT:  FOSS, Margit
;    APPLICANT:  McNALLY, Francis J
;    APPLICANT:  LAURENSON, Patricia
;    APPLICANT:  HERSKOWITZ, Ira
;    APPLICANT:  LI, Joachim J
```

```
;    APPLICANT:  GAVIN, Kimberly
;    TITLE OF INVENTION:  ORIGIN OF REPLICATION COMPLEX GENES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,105
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59032/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  430 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-484-105-22

  Query Match             4.7%;  Score 105;  DB 1;  Length 430;
  Best Local Similarity   21.9%;  Pred. No. 0.013;
  Matches   61;  Conservative   42;  Mismatches   87;  Indels   88;  Gaps   13;

Qy      85 PKKTGSRPVVLLQHGLVGGASNWISNLPNN------SLGF-----ILADAGFDVW--MGN 131
           |:| |||           |  |   || :|        :| :  : :  |
Db      25 PEKEGSR------QKKTNGKENASRNLQSNLEEDLEQLGFEDETVSMAQSAIENYFMQGK 78

Qy     132 S----RGNAWSRKHK------TLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIY 181
           |      || ||: :      |  |::|:  : :  :  : ||| : |:| :|   |
Db      79 SASERMNNAKSRRGRRAGNGNTEEIEEDDEISNAITDFTKCDLPGLRNYITKKDNTEFEK 138

Qy     182 YVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFG 241
            : :     |           | |:| |                    :::: |:
Db     139 RLEHLADNDFG-----------KWKLYLAAG-----------------FNILLHGVGS 168

Qy     242 KKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGFNTN---NMNMSRASVYAAHTLA 298
           |:: |  | |  :|  |   :|   :   |: :|||  | |   |:|: |
Db     169 KRDVL--TEFENELSDYTYMRVDARKDGLNVKVLLGAINENMKLNCNVKRG--------- 217

Qy     299 GTSVQNILHWSQAV----NSGEL-------RAFDWGSE 325
               |: : |::::    || :|        | || |:
Db     218 ----QSTISWARSIRRKMNSQQLILIIDNIEAPDWRSD 251


RESULT  12
US-08-484-106-22
; Sequence 22, Application US/08484106
; Patent No. 5514618
;  GENERAL INFORMATION:
;    APPLICANT:  ZEILLMAN, Bruce
;    APPLICANT:  BELL, Stephen P
;    APPLICANT:  KOBAYASHI, Ryuji
;    APPLICANT:  RINE, Jasper
;    APPLICANT:  ROSS, Margit
;    APPLICANT:  McNALLY, Francis J
;    APPLICANT:  LAURENSON, Patricia

;    APPLICANT:  HERSKOWITZ, Ira
;    APPLICANT:  LI, Joachim J
;    APPLICANT:  GAVIN, Kimberly
;    TITLE OF INVENTION:  ORIGIN OF REPLICATION COMPLEX GENES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,106
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59032/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  430 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-484-106-22

  Query Match             4.7%;  Score 105;  DB 1;  Length 430;
  Best Local Similarity   21.9%;  Pred. No. 0.013;
  Matches   61;  Conservative   42;  Mismatches   87;  Indels   88;  Gaps   13;

Qy      85 PKKTGSRPVVLLQHGLVGGASNWISNLPNN------SLGF-----ILADAGFDVW--MGN 131
           |:| |||           |  |   || :|        :| :  : :  |
Db      25 PEKEGSR------QKKTNGKENASRNLQSNLEEDLEQLGFEDETVSMAQSAIENYFMQGK 78

Qy     132 S----RGNAWSRKHK------TLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIY 181
           |      || ||: :      |  |::|:  : :  :  : ||| : |:| :|   |
Db      79 SASERMNNAKSRRGRRAGNGNTEEIEEDDEISNAITDFTKCDLPGLRNYITKKDNTEFEK 138

Qy     182 YVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFG 241
            : :     |           | |:| |                    :::: |:
Db     139 RLEHLADNDFG-----------KWKLYLAAG-----------------FNILLHGVGS 168

Qy     242 KKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGFNTN---NMNMSRASVYAAHTLA 298
           |:: |  | |  :|  |   :|   :   |: :|||  | |   |:|: |
Db     169 KRDVL--TEFENELSDYTYMRVDARKDGLNVKVLLGAINENMKLNCNVKRG--------- 217

Qy     299 GTSVQNILHWSQAV----NSGEL-------RAFDWGSE 325
               |: : |::::    || :|        | || |:
Db     218 ----QSTISWARSIRRKMNSQQLILIIDNIEAPDWRSD 251


RESULT  13
US-08-602-359A-34
; Sequence 34, Application US/08602359A
; Patent No. 5942430
;  GENERAL INFORMATION:
;    APPLICANT:  ROBERTSON, Daniel E.
;    APPLICANT:  MURPHY, Dennis
;    APPLICANT:  REID, John
;    APPLICANT:  MAFFIA, Anthony
;    APPLICANT:  LINK, Steven
```

```
;    APPLICANT:  SWANSON, Ronald V.
;    APPLICANT:  WARREN, Patrick V.
;    APPLICANT:  KOSMOTKA, Anna
;    TITLE OF INVENTION:  ESTERASES
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & RICHARDSON P.C.
;      STREET:  4225 EXECUTIVE SQUARE, STE 1400
;      CITY:  LA JOLLA
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/602,359A
;      FILING DATE:  February 16, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HAILE, LISA A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09010/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-678-5070
;      TELEFAX:  619-678-5099
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  346 AMINO ACIDS
;      TYPE:  AMINO ACID
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-08-602-359A-34


  Query Match             4.5%;  Score 102;  DB 2;  Length 346;
  Best Local Similarity   22.8%;  Pred. No. 0.019;
  Matches   41;  Conservative   33;  Mismatches   74;  Indels   32;  Gaps    6;

Qy      20 LILVAYMFQRNVNSVHMPT----KAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           |:| |       | || :|      :: ||:   ::     :|   :|: |  ||| ::
Db      60 LLLAAVAVLALVVSVVVPAYVAYSSLHPESCRPVAP----EGLTYKEFSVTAEDGLVV-- 113

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVG-GASNWISNLPNNSLGFILADAGFDVWMGNSRG 134
             || |      |  || :| ||  |  :: ::: |        | : |: | : : ||
Db     114 ----RGWVLGPGAGGNPVFVLMHGYTGCRSAPYMAVLARE-----LVEWGYPVVVFDFRG 164

Qy     135 NAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFI 194
           :  |    |:                  |  ||: :: ::    :|  ||:| |  :  :
Db     165 HGESGGSTTIG-----------PREVLDARAVVGYVSERFPGRRIILVGFSMGGAVAIV 212


RESULT  14
US-08-232-519-2
; Sequence 2, Application US/08232519
; Patent No. 5484725
;  GENERAL INFORMATION:
;    APPLICANT:  Kageyama, Bunji
;    APPLICANT:  Nakae, Masanori
;    APPLICANT:  Yagi, Shigeo
;    TITLE OF INVENTION:  NORBORANE TYPE ESTER HYDROLASE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Neave
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,519
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 5-96286
;      FILING DATE:  22-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley Jr, James F
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  SHGN-4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 596-9000
;      TELEFAX:  (212) 596-9090
;      TELEX:  14-8367
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  388 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-232-519-2


  Query Match             4.4%;  Score 98.5;  DB 1;  Length 388;
  Best Local Similarity   20.2%;  Pred. No. 0.052;
  Matches   81;  Conservative   64;  Mismatches  142;  Indels  113;  Gaps   20;

Qy      38 TKAVDPEAFMNISEIIQHQGYPCEEYE--VATEDGYILSVNRIPRG--LVQPKKTGSR-- 91
           |  || |||:         : :  |  :   ||||       ::  |   :  :  |::
Db      84 TPDVDIEAFI--------KNFHAENPQDAAATEDASAGEPKQVTVGEHTLNVRDVGTQEG 135

Qy      92 -PVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKH---KTLSID 147
            |:||: ||  |  |||:  |  ::|        ||:       |:  | |:    ||:
Db     136 TPIVLV-HGFGGDISNWL--LTQDALAAERRVIAFDL-----PGHGASSKNVGTGTLAF- 186

Qy     148 QDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIK 207
                          |  |:: :||    || : ||:| |   | || : : :   ::
Db     187 ---------------LAGVVSELLQTLKIEKAHVVGHSLG---GGIALTLLRDHPDQVA 227

Qy     208 MYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIY---LCGQVI 264
               |||       |     |:          |   |     : : |:::|   | |: :
Db     228 SLNLLAPAGL---GKDVNADFI--------SAFVDSESSRDMKAVLQMLVYNKALVGRKM 276

Qy     265 LDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGS 324
           :| :                :: :|    | | :|      |           : :| :   |:
Db     277 VDAV----------LRARRLDGARD---ALHVIAKACFPN------GHQADDLHSVLAGA 317

Qy     325 ETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHK--NIP 382
           ||                      || :: | :| : :  :    |   :   :| :  ::|
Db     318 ET--------------------PTQIFWGKEDEILSVSNAAGLPDVIPVTVYEETGHLP 356

Qy     383 EWAHVDFIWGLDAPHRMYNEIIHL-MQQEETNLSQGRCEA 421
           :  |   :           |: | | ::  |  ||  | :|
Db     357 QLEHATDV----------NKAIALFVKDPEAALSMARMDA 386


RESULT  15
US-08-456-956-2
; Sequence 2, Application US/08456956
; Patent No. 5686285
;  GENERAL INFORMATION:
;    APPLICANT:  Kageyama, Bunji
;    APPLICANT:  Nakae, Masanori
;    APPLICANT:  Yagi, Shigeo
;    TITLE OF INVENTION:  No. 5686285bornane Type Ester Hydrolase
;    NUMBER OF SEQUENCES:  2
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Neave
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/456,956
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 5-96286
;      FILING DATE:  22-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/232,519
;      FILING DATE:  22-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley Jr, James F
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  SHGN-4 DIV 1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)596-9000
;      TELEFAX:  (212)596-9090
;      TELEX:  14-8367
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  388 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-456-956-2


  Query Match             4.4%;  Score 98.5;  DB 1;  Length 388;
  Best Local Similarity   20.2%;  Pred. No. 0.052;
  Matches   81;  Conservative   64;  Mismatches  142;  Indels  113;  Gaps   20;

Qy      38 TKAVDPEAFMNISEIIQHQGYPCEEYE--VATEDGYILSVNRIPRG--LVQPKKTGSR-- 91
           |  || |||:        : :  |  :   ||||       ::  |   :  :  |::
Db      84 TPDVDIEAFI--------KNFHAENPQDAAATEDASAGEPKQVTVGEHTLNVRDVGTQEG 135

Qy      92 -PVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKH---KTLSID 147
            |:||: ||  |  |||:  |  ::|        ||:      |:  | |:    ||:
Db     136 TPIVLV-HGFGGDISNWL--LTQDALAAERRVIAFDL-----PGHGASSKNVGTGTLAF- 186

Qy     148 QDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIK 207
                       |  |:: :||    || : ||:| |   | || : : :   ::
Db     187 ---------------LAGVVSELLQTLKIEKAHVVGHSLG---GGIALTLLRDHPDQVA 227

Qy     208 MYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIY---LCGQVI 264
               |||       |     |:          |   |     : : |:::|   | |: :
Db     228 SLNLLAPAGL---GKDVNADFI--------SAFVDSESSRDMKAVLQMLVYNKALVGRKM 276

Qy     265 LDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGS 324
           :| :                 :: :|    | | :|     |          : :| :   |:
Db     277 VDAV----------LRARRLDGARD---ALHVIAKACFPN------GHQADDLHSVLAGA 317

Qy     325 ETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHK--NIP 382
           ||                    || :: | :| : :  :   |   :   :| :  ::|
Db     318 ET--------------------PTQIFWGKEDEILSVSNAAGLPDVIPVTVYEETGHLP 356

Qy     383 EWAHVDFIWGLDAPHRMYNEIIHL-MQQEETNLSQGRCEA 421
           :  |   :          |: | | ::  |  ||  | :|
Db     357 QLEHATDV----------NKAIALFVKDPEAALSMARMDA 386
```

```
Search completed: January 30, 2002, 23:01:20
Job time: 3001 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 30, 2002, 22:13:59 ; Search time 57.04 Seconds
(without alignments)
564.899 Million cell updates/sec

Title: US-09-578-063-47
Perfect score: 2247
Sequence: 1 MLETLSRQWIVSHRMEMWLL.....IHLMQQEETNLSQGRCEAVL 423

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1290	57.4	399	2	S41408	lysosomal acid lip
2	1290	57.4	399	2	G01416	lysosomal acid lip
3	1164.5	51.8	395	1	LIRTT	triacylglycerol li
4	1139	50.7	397	1	JC4017	triacylglycerol li
5	1119	49.8	398	2	S07145	triacylglycerol li
6	752.5	33.5	403	2	T33198	hypothetical prote
7	751	33.4	411	2	T22290	hypothetical prote
8	730	32.5	405	2	T22675	hypothetical prote
9	655.5	29.2	426	2	T20480	hypothetical prote
10	574.5	25.6	559	2	JT0949	egg-specific prote
11	505	22.5	443	2	T39540	triglyceride lipas
12	478.5	21.3	1585	2	T31611	hypothetical prote
13	475	21.1	467	2	T41053	triglyceride lipas
14	445	19.8	344	2	E84526	probable lysosomal
15	422.5	18.8	548	2	S37969	probable triacylgl
16	412.5	18.4	460	2	T39443	probable triglycer
17	411.5	18.3	413	2	T43170	probable triacylgl
18	294	13.1	431	2	D86318	protein F15H18.6 [
19	285	12.7	538	2	S64842	probable membrane
20	271.5	12.1	573	2	S64754	probable membrane
21	234.5	10.4	509	2	G96766	protein lipase F2P
22	149	6.6	62	2	S59904	lipase, pregastric
23	143	6.4	336	2	F83425	probable esterase/
24	126	5.6	987	2	A64474	hypothetical prote
25	113	5.0	412	2	B72391	conserved hypothet
26	112.5	5.0	40	2	S19539	triacylglycerol li
27	105	4.7	430	2	T23001	hypothetical prote
28	103.5	4.6	798	2	F96714	probable protease
29	101.5	4.5	929	2	JH0262	QutR protein - Eme
30	101	4.5	436	2	JC5021	platelet-activatin
31	99.5	4.4	286	2	D75217	probable 2-acetyl-
32	99.5	4.4	1280	2	B34087	hypothetical prote
33	99	4.4	460	2	G96764	unknown protein F2
34	97.5	4.3	322	2	B70957	probable ephA prot
35	97.5	4.3	456	2	E69391	hypothetical prote
36	96.5	4.3	287	1	E64053	hypothetical prote
37	96.5	4.3	1275	2	I38588	reverse transcript
38	96	4.3	809	2	S67153	probable membrane
39	96	4.3	934	2	B82364	DNA polymerase I V
40	95.5	4.3	1259	4	GNHUL1	retrovirus-related
41	95	4.2	973	2	T41201	isoleucyl-trna syn
42	94.5	4.2	561	2	G83913	hypothetical prote
43	93	4.1	1275	2	B28096	line-1 protein ORF
44	93	4.1	2311	1	TVCHSR	kinase-related pro
45	92.5	4.1	1275	2	S65824	reverse transcript

ALIGNMENTS

RESULT 1
S41408
lysosomal acid lipase (EC 3.1.1.-) / sterol esterase (EC 3.1.1.13) precursor - human
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 18-Jun-1999
C;Accession: S41408; A39315; S47187
R;Ameis, D.; Merkel, M.; Eckerskorn, C.; Greten, H.
Eur. J. Biochem. 219, 905-914, 1994
A;Title: Purification, characterization and molecular cloning of human hepatic lysoso
A;Reference number: S41408; MUID:94155897
A;Accession: S41408
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-399 <AME>
A;Cross-references: EMBL:X76488; NID:g434305; PIDN:CAA54026.1; PID:g434306
R;Anderson, R.A.; Sando, G.N.
J. Biol. Chem. 266, 22479-22484, 1991
A;Title: Cloning and expression of cDNA encoding human lysosomal acid lipase/choleste
A;Reference number: A39315; MUID:92042192
A;Accession: A39315
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-15,'P',17-399 <AND>
A;Cross-references: GB:M74775; NID:g187151; PIDN:AAA59519.1; PID:g187152
R;Du, H.; Gregory, G.A.
submitted to the EMBL Data Library, April 1994
A;Description: Structural conservation of putative functional motifs between mouse an
A;Reference number: S47187
A;Accession: S47187
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-22,'R',24-399 <DUH>
A;Cross-references: EMBL:Z31690; NID:g506430; PIDN:CAA83495.1; PID:g506431
C;Genetics:
A;Gene: GDB:LIPA
A;Cross-references: GDB:120153; OMIM:278000
A;Map position: 10q24-10q25
C;Superfamily: triacylglycerol lipase, lingual
C;Keywords: carboxylic ester hydrolase; glycoprotein

Query Match 57.4%; Score 1290; DB 2; Length 399;
Best Local Similarity 61.0%; Pred. No. 6.1e-106;
Matches 242; Conservative 59; Mismatches 94; Indels 2; Gaps 1;

```
Qy      15 MEMWLLILVAYMFQRNVNSVHMPTK--AVDPEAFMNISEIIQHQGYPCEEYEVATEDGYI 72
           |:|  | ||  :   ::|    |  ||||| ||:|||| : |:| ||| | ||||||
Db       1 MKMRFLGLVVCLVLWTLHSEGSGGKLTAVDPETNMNVSEIISYWGFPSEEYLVETEDGYI 60

Qy      73 LSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNS 132
           | :|||| |       | :||| |||||:  :|||::|| |:|||||||||||||||||
Db      61 LCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANSSLGFILADAGFDVWMGNS 120
```

```
Qy     133 RGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMG 192
           ||| |||||||||: |||||||||||||::|||| ||||| |||||::||||:|||||:|
Db     121 RGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFILNKTGQEQVYYVGHSQGTTIG 180

Qy     193 FIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFL 252
           ||||| :||||::|||:||| |:|:|    ||  |   ||| :|| ||| |||| |: ||
Db     181 FIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDLFGDKEFLPQSAFL 240

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
           : |  ::|  ||| ::| |:  || |||  |:|||  ||  |: ||||||:|||||||
Db     241 KWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSRVDVYTTHSPAGTSVQNMLHWSQAV 300

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
             : :||||||  ||    ||  |  | |:|| ||||:|:|| |||::  || :||:::
Db     301 KFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDMLVPTAVWSGGHDWLADVYDVNILLTQI 360

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQ 409
           |||::|::|||| |:|||||||| |:||:||:||::
Db     361 TNLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRK 397


RESULT   2
G01416
lysosomal acid lipase - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 18-Jun-1999
C;Accession: G01416
R;Du, H.
submitted to the EMBL Data Library, April 1994
A;Reference number: G06919
A;Accession: G01416
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-399 <DUX>
A;Cross-references: EMBL:U08464; NID:g505052; PIDN:AAB60328.1; PID:g505053
C;Superfamily: triacylglycerol lipase, lingual
C;Keywords: glycoprotein


  Query Match             57.4%;  Score 1290;  DB 2;  Length 399;
  Best Local Similarity   61.0%;  Pred. No. 6.1e-106;
  Matches  242;  Conservative   59;  Mismatches   94;  Indels    2;  Gaps    1;

Qy      15 MEMWLLILVAYMFQRNVNSVHMPTK--AVDPEAFMNISEIIQHQGYPCEEYEVATEDGYI 72
           |:|  | ||  :    ::|     |  ||||| ||:|||| : |:| ||| | ||||||
Db       1 MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYWGFPSEEYLVETEDGYI 60

Qy      73 LSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNS 132
           | :|||| |       | :||| |||||:  :|||::|| |:||||||||||||||||||
Db      61 LCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANSSLGFILADAGFDVWMGNS 120

Qy     133 RGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMG 192
           ||| |||||||||: |||||||||||||::|||| ||||| |||||::||||:|||||:|
Db     121 RGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFILNKTGQEQVYYVGHSQGTTIG 180

Qy     193 FIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFL 252
           ||||| :||||::|||:||| |:|:|    ||  |   ||| :|| ||| |||| |: ||
Db     181 FIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDLFGDKEFLPQSAFL 240

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
           : |  ::|  ||| ::| |:  || |||  |:|||  ||  |: ||||||:|||||||
Db     241 KWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSRVDVYTTHSPAGTSVQNMLHWSQAV 300

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
             : :||||||  ||    ||  |  | |:|| ||||:|:|| |||::  || :||:::
Db     301 KFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDMLVPTAVWSGGHDWLADVYDVNILLTQI 360

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQ 409
           |||::|::|||| |:|||||||| |:||:||:||::
Db     361 TNLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRK 397


RESULT   3
LIRTT
triacylglycerol lipase (EC 3.1.1.3) precursor, lingual - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 11-Jun-1999
C;Accession: A23045
R;Docherty, A.J.P.; Bodmer, M.W.; Angal, S.; Verger, R.; Riviere, C.; Lowe, P.A.; L
Nucleic Acids Res. 13, 1891-1903, 1985
A;Title: Molecular cloning and nucleotide sequence of rat lingual lipase cDNA.
A;Reference number: A23045; MUID:85215587
A;Accession: A23045
A;Molecule type: mRNA
A;Residues: 1-395 <DOC>
A;Cross-references: GB:X02309; NID:g56595; PIDN:CAA26179.1; PID:g56596
A;Experimental source: strain Sprague-Dawley
A;Note: the partial sequence of the mature protein from a different, unspecified st
C;Comment: This acid-stable lipase is secreted by the serous (von Ebner's) glands a
C;Superfamily: triacylglycerol lipase, lingual
C;Keywords: carboxylic ester hydrolase; glycoprotein; lipid digestion; saliva; sero
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-395/Product: triacylglycerol lipase, lingual #status predicted <MPT>
F;33,68,98,184,270/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             51.8%;  Score 1164.5;  DB 1;  Length 395;
  Best Local Similarity   54.9%;  Pred. No. 7.1e-95;
  Matches  217;  Conservative   69;  Mismatches  108;  Indels    1;  Gaps    1;

Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||||:: : :      : :       :||| ||||::| : ||||:|||| |||||| |
Db       1 MWLLLITSVISTFGGAHGLFGKLGPGNPEANMNISQMITYWGYPCQEYEVVTEDGYILGV 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |    :  | |||| |||||:  |:|||:|||||| |:|||||:|||:||||||
Db      61 YRIPHGKNNSENIGKRPVVYLQHGLIASATNWIANLPNNSLAFMLADAGYDVWLGNSRGN 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           |||||:   | |  ||||||:||||::|||4 ||||:||||||||:|||:|||||:||||
Db     121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATINFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQL 255
           ||| | ||:||| ::|||||:|||| :||  |   :|  : | :|||| ||  | |   |
Db     181 FSTNPTLAKKIKTFYALAPVATVKYTQSPLKKISFIPTFLFKLMFGKKMFLPHTYFDDFL 240

Qy     256 VIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSG 315
           :| : :|| :||| : :  ||:  |:|:|| ||  |  ||||||: |||:| | ||
Db     241 GTEVCSREVLDLLCSNTLFIFCGFDKKNLNVSRFDVYLGHNPAGTSVQDFLHWAQLVRSG 300

Qy     316 ELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNL 375
           : :||:||| ::|:   || ||  | |  |||| |:| || | |::|:|| ||| :::||
Db     301 KFQAFNWGSPSQNMLHYNQKTPPEYDVSAMTVPVAVWNGGNDILADPQDVAMLLPKLSNL 360

Qy     376 IYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ::|| |  : |:|||| :|||  :|||:| :| ::
Db     361 LFHKEILAYNHLDFIWAMDAPQEVYNEMISMMAED 395


RESULT   4
JC4017
triacylglycerol lipase (EC 3.1.1.3) PGE precursor - bovine
N;Alternate names: pregastric esterase
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: JC4017; S64678
R;Timmermans, M.Y.J.; Teuchy, H.; Kupers, L.P.M.
Gene 147, 259-262, 1994
A;Title: The cDNA sequence encoding bovine pregastric esterase.
A;Reference number: JC4017; MUID:95011625
A;Accession: JC4017
A;Molecule type: mRNA
A;Residues: 1-397 <TIM>
A;Cross-references: GB:L26319; NID:g600756; PIDN:AAA57037.1; PID:g600757
A;Experimental source: tongue
```

```
R;Timmermans, M.Y.J.; Reekmans, G.; Teuchy, H.J.H.; Kupers, L.P.M.
Biochem. J. 314, 931-936, 1996
A;Title: Inhibition studies on calf pregastric esterase: the enzyme has no functional th
A;Reference number: S64678; MUID:96177869
A;Accession: S64678
A;Molecule type: protein
A;Residues: 24-40;248-253 <TIW>
C;Comment: Pregastric esterase is a major fat-digesting enzyme.
C;Genetics:
A;Gene: pge
C;Superfamily: triacylglycerol lipase, lingual
C;Keywords: blocked amino end; carboxylic ester hydrolase; glycoprotein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-397/Product: pregastric esterase #status predicted <MAT>
F;33,270,326/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;171/Active site: Ser #status predicted


  Query Match            50.7%;  Score 1139;  DB 1;  Length 397;
  Best Local Similarity  53.7%;  Pred. No. 1.3e-92;
  Matches  212;  Conservative   68;  Mismatches  113;  Indels    2;  Gaps    2;

Qy      18 WLLILVAYMFQRNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNR 77
           |||: | ::          :   | :||| ||:|::| : ||| | ::| | ||||| | |
Db       3 WLLVTVCFIHMSGNAFCFLGKIAKNPEASMNVSQMISYWGYPSEMHKVITADGYILQVYR 62

Qy      78 IPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAW 137
           || |       | |||| ||||:| |:|||||| ||||:|||||:|||:|||||| |
Db      63 IPHGKNNANHLGQRPVVFLQHGLLGSATNWISNLPKNSLGFLLADAGYDVWLGNSRGNTW 122

Qy     138 SRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFS 197
           :::|   | |  ||||||:|||| :|||: |:|||::|||:|::|||:|||||:||||||
Db     123 AQEHLYYSPDSPEFWAFSFDEMAEYDLPSTIDFILRRTGQKKLHYVGHSQGTTIGFIAFS 182

Qy     198 TMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LV 256
           | | ||:|||:::||||:|||| ||   |  |:|  : | :|| | | |   || | |
Db     183 TSPTLAEKIKVFYALAPVATVKYTKSLFNKLALIPHFLFKIIFGDKMF-YPHTFLEQFLG 241

Qy     257 IYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGE 316
           : :| :  || :| | :  : | :  | ||||  || ||  |||||||| ||| ||| ||:
Db     242 VEMCSRETLDVLCKNALFAITGVDNKNFNMSRLDVYIAHNPAGTSVQNTLHWRQAVKSGK 301

Qy     317 LRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLI 376
           :|||||:  :||   :|||| | :  | || |:|:   | |::|:||  |||:::|||
Db     302 FQAFDWGAPYQNLMHYHQPTPPIYNLTAMNVPIAVWSADNDLLADPQDVDFLLSKLSNLI 361

Qy     377 YHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQEE 411
           ||| || : |:|||| :|||  :||||: || :::
Db     362 YHKEIPNYNHLDFIWAMDAPQEVYNEIVSLMAEDK 396


RESULT   5
S07145
triacylglycerol lipase (EC 3.1.1.3) precursor, gastric - human
C;Species: Homo sapiens (man)
C;Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 18-Jun-1999
C;Accession: S07145; S27102; S04942
R;Bodmer, M.W.; Angal, S.; Yarranton, G.T.; Harris, T.J.R.; Lyons, A.; King, D.J.; Pierc
Biochim. Biophys. Acta 909, 237-244, 1987
A;Title: Molecular cloning of a human gastric lipase and expression of the enzyme in yea
A;Reference number: S07145; MUID:87299724
A;Accession: S07145
A;Molecule type: mRNA
A;Residues: 1-398 <BOD1>
A;Cross-references: EMBL:X05997; NID:g31771; PIDN:CAA29413.1; PID:g758063
A;Accession: S27102
A;Molecule type: protein
A;Residues: 20-24,'X',26,'X',28-33,'X',35-42,'X',44,'X',46-48,'X',50,'X',52,'X',54-56 <E
A;Note: it is uncertain whether Met-1 or Met-7 is the intiator
R;Bernbaeck, S.; Blaeckberg, L.
Eur. J. Biochem. 182, 495-499, 1989
A;Title: Human gastric lipase. The N-terminal tetrapeptide is essential for lipid bindin
A;Reference number: S04942; MUID:89325292
A;Accession: S04942
A;Molecule type: protein
A;Residues: 20-45 <BER>
C;Superfamily: triacylglycerol lipase, lingual
C;Keywords: carboxylic ester hydrolase; extracellular protein; glycoprotein; lipid di
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-398/Product: triacylglycerol lipase, gastric #status experimental <MAT>
F;34,99,185,271/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match            49.8%;  Score 1119;  DB 2;  Length 398;
  Best Local Similarity  52.2%;  Pred. No. 7.5e-91;
  Matches  210;  Conservative   73;  Mismatches  107;  Indels   12;  Gaps    2;

Qy      17 MWLLILVAYMFQRNVNSVHMPTKAV-------DPEAFMNISEIIQHQGYPCEEYEVATED 69
           ||||: :| :      ||   | :         ||  ||||::| : ||| ||||| |||
Db       1 MWLLLTMASLI-----SVLGTTHGLFGKLHPGSPEVTMNISQMITYWGYPNEEYEVVTED 55

Qy      70 GYILSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWM 129
           |||| ||||| |       || |||| |||||:  |:||||||||||| |||||||:|||:
Db      56 GYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSLAFILADAGYDVWL 115

Qy     130 GNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGT 189
           |||||| |:|::   | |  ||||||:||||::|||| |:||::||||::::|||:||||
Db     116 GNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKKTGQKQLHYVGHSQGT 175

Qy     190 TMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQT 249
           |:||||||| | ||::|| ::||||:||||: ||   |.  :|  : | :|| | |
Db     176 TIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVPQSLFKFIFGDKIFYPHN 235

Qy     250 RFLRQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWS 309
           | : |   :| : :|: :||| : :: ||:: | | ||  || :|  |||||||: ||:
Db     236 FFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLDVYLSHNPAGTSVQNMFHWT 295

Qy     310 QAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLL 369
           ||| ||: :|:|||| :|    :|  |  |:|  | || |:| ||:| |::|:|| :||
Db     296 QAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVPIAVWNGGKDLLADPQDVGLLL 355

Qy     370 SEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQEE 411
           :: |||||| || : |:|||| :|||  :||:|: :: :::
Db     356 PKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSMISEDK 397


RESULT   6
T33198
hypothetical protein ZK6.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C;Accession: T33198
R;Wu, X.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid ZK6.
A;Reference number: Z21301
A;Accession: T33198
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-403 <WUX>
A;Cross-references: EMBL:AF067942; PIDN:AAC17694.1; GSPDB:GN00023; CESP:ZK6.7
A;Experimental source: strain Bristol N2; clone ZK6
C;Genetics:
A;Gene: CESP:ZK6.7
A;Map position: 5
A;Introns: 29/3; 63/1; 219/3; 319/3; 365/3
C;Superfamily: triacylglycerol lipase, lingual


  Query Match            33.5%;  Score 752.5;  DB 2;  Length 403;
  Best Local Similarity  41.1%;  Pred. No. 1.7e-58;
  Matches  168;  Conservative   66;  Mismatches  142;  Indels   33;  Gaps   12;

Qy      17 MW-LLILVAYMFQRNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||   : :|  | ::|   |      |||  |   :||:  |||   | |||:||||| :
Db       1 MWRFAVFLAAFFVQDVVGSH-----GDPELHMTTPQIIERWGYPAMIYTVATDDGYILEM 55
```

```
Qy      76 NRIPRGLVQPKKT------GSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWM 129
           :||| |   ||      | |||| :||||:  :|:|: |||: | ||:|||||||||:
Db      56 HRIPFG-----KTNVTWPNGKRPVVFMQHGLLCASSDWVVNLPDQSAGFLFADAGFDVWL 110

Qy     130 GNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGT 189
           || ||| :| ||| |       || :|:|||| :|| |:|| :|: ||| :||:|:|||
Db     111 GNMRGNTYSMKHKDLKPSHSAFWDWSWDEMATYDLNAMINHVLEVTGQDSVYYMGHSQGT 170

Qy     190 TMGFIAFSTMP-ELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKG---LFGKKEF 245
               |   |      |:||| :|||||| :||| |   : |     :   |   :||  ||
Db     171 LTMFSHLSKDDGSFAKKIKKFFALAPIGSVKHIKGFLSFFANYFSLEFDGWFDIFGAGEF 230

Qy     246 LYQTRFLRQLVIYLCGQVILD-QICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQN 304
           |     ::     :|| : ::  :| |:: |: |  ::  | :|  ||| |  |||| ||
Db     231 LPNNWAMKLAAKDICGGLKVEADLCDNVLFLIAGPESDQWNQTRVPVYATHDPAGTSTQN 290

Qy     305 ILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMT----VPTAMWTGGQDWLS 360
           |:|| | |: | : |:|||::| | :|  |  |  |  | |        ::    |||:
Db     291 IVHWMQMVHHGGVPAYDWGTKT-NKKKYGQANPPEY---DFTAIKGTKIYLYWSDADWLA 346

Qy     361 NPEDV-KMLLSEVTNLIYHKN--IPEWAHVDFIWGLDAPHRMYNEIIHL 406
           :  ||   ||: :   |  :|  :|:: |:|| ||| ||  :|   | |
Db     347 DTPDVPDYLLTRLNPAIVAQNNHLPDYNHLDFTWGLRAPDDIYRPAIKL 395
```

RESULT 7
T22290
hypothetical protein F46B6.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T22290
R;McMurray, A.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z19542
A;Accession: T22290
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-411 <WIL>
A;Cross-references: EMBL:Z70780; PIDN:CAA94824.1; GSPDB:GN00023; CESP:F46B6.8
A;Experimental source: clone F46B6
C;Genetics:
A;Gene: CESP:F46B6.8
A;Map position: 5
A;Introns: 35/3; 69/1; 104/1; 226/3
C;Superfamily: triacylglycerol lipase, lingual

Query Match 33.4%; Score 751; DB 2; Length 411;
Best Local Similarity 40.4%; Pred. No. 2.3e-58;
Matches 163; Conservative 72; Mismatches 138; Indels 30; Gaps 12;

```
Qy      20 LILVAYMFQRNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIP 79
           ||  :::  : :|: |      |||  || |:||:  ||  | : | |||||||| : |||
Db      10 LICTSFLLLQTINA-HPDD---DPELNMNTSQIIERWGYKAEVHTVTTEDGYILQMQRIP 65

Qy      80 RGLVQPKKT------GSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSR 133
             |     ||      | |||:||||||:  ||:|: |||  |  |: |||||||:|| |
Db      66 YG-----KTSVTWPNGKRPVILLQHGLLACASDWVDNLPTQSAAFVFADAGFDVWLGNVR 120

Qy     134 GNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGF 193
           |  : ||:  |   :  || ||:|||||::|:||::: :|  |||| :||:|:|||| : |
Db     121 GTTYGRKNTKLDPSETAFWQFSWDEMAQYDVPAMVDHVLAMTGQENLYYMGHSQGTLIMF 180

Qy     194 --IAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKG---LFGKKEFLYQ 248
             :|  |    |:||| |||||||  ||: |   : |          |   ||| |:||
Db     181 THLAKDTDGSFAKKIKRYFALAPIGAVKNIKGFLSYFAHKFSPEFDGWYELFGSKDFLPD 240

Qy     249 TRFLRQLVIYLCGQVILD-QICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILH 307
             :     :||    : ::| | : |:||  ::  | || ::|::   |||| |||:|
Db     241 NWITKMAAKDICGASEKEAELCDNELFLIGGPESDQWNASRTAIYSSQDPAGTSTQNIVH 300

Qy     308 WSQAVNSGELRAFDWGSETKNLEKCNQPTPVRY---RVRDMTVPTAMWTGGQDWLSNPED 364
           | | | :| : ||||| :  | :| .| ||  |    ::   :  ::    ||| :| |
Db     301 WMQMVRNGRVPAFDWGKKI-NKKKYGQDTPPEYDFGAIKGTKI--HLYWSDDDWLGDPTD 357

Qy     365 V-KMLLSEVTNLIYHKNI--PEWAHVDFIWGLDAPHRMYNEII 404
           :   || |:   :  :|:   |: |:|| |||:|  ::|:: |
Db     358 IHDFLLKELNPAVIAENVNLKEYNHLDFTWGLNATFQIYDKAI 400
```

RESULT 8
T22675
hypothetical protein F54F3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T22675
R;Percy, C.; Lloyd, C.
submitted to the EMBL Data Library, September 1996
A;Reference number: Z19598
A;Accession: T22675
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-405 <WIL>
A;Cross-references: EMBL:Z79696; PIDN:CAB01973.1; GSPDB:GN00023; CESP:F54F3.3
A;Experimental source: clone F54F3
C;Genetics:
A;Gene: CESP:F54F3.3
A;Map position: 5
A;Introns: 31/3; 65/1; 274/3; 366/3
C;Superfamily: triacylglycerol lipase, lingual

Query Match 32.5%; Score 730; DB 2; Length 405;
Best Local Similarity 38.6%; Pred. No. 1.6e-56;
Matches 161; Conservative 69; Mismatches 155; Indels 32; Gaps 11;

```
Qy      15 MEMWLLILVAYMFQRNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILS 74
           |  |  :::| :       |      |||  |   :||   |||   |:| |||||||
Db       1 MRSWSTVMLAVLATAATVFGH----DADPEMKMTTPQIIMRWGYPAMIYDVTTEDGYILE 56

Qy      75 VNRIPRGLVQPKKT------GSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVW 128
           ::||| |     ||      | :||| :||||   :|||: |||  |  |: ||||:|||
Db      57 LHRIPYG-----KTNVTWPNGKKPVVFMQHGLECSSSNWVVNLPTESAAFLFADAGYDVW 111

Qy     129 MGNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQG 188
           :|| ||| :| ||| |       || :|:||| ::||||:|   |: |||: :||:|:|||
Db     112 LGNFRGNTYSMKHKNLKPSHSAFWDWSWDEMQQYDLPAMIEKALEVTGQDSLYYIGHSQG 171

Qy     189 T-TMGFIAFSTMPE----LAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKG---LF 240
           | ||    || : |       ||| :|||||: :||| |     |     :   |   :|
Db     172 TLTM----FSRLSEDKVGWGNKIKKFFALAPVGSVKHIKGALKFFADYFSLEFDGWFDVF 227

Qy     241 GKKEFLYQTRFLRQLVIYLC-GQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAG 299
           |  |||     :: :   :| |  :   :| ::| |: | :| :| :|  :| ||| ||
Db     228 GSGEFLPNNWIMKLVSESVCAGLKVEAGVCDDVMFLIAGPESNQLNATRVPIYVAHTPAG 287

Qy     300 TSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWL 359
           || |||:|| | |   :|:| |  | :   |     |   :  | :: |  |||
Db     288 TSTQNIVHWIQMVRHGGTPKYDYG-EKGNKKHYGQANVPAYDFTTVNRPVYLYWGDSDWL 346

Qy     360 SNPEDV-KMLLSEV--TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQEETN 413
           ::| ||   ||: :  : :: :  : :: |:|||||| ||  :|  || ::: :  |
Db     347 ADPTDVTDFLLTHLNPSTVVQNNKLIDYNHLDFIWGLRAPKDIYEPIIDIVRNDVLN 403
```

RESULT 9
T20480
hypothetical protein F01G10.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T20480
R;Hembry, C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19280
A;Accession: T20480

```
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-426 <WIL>
A;Cross-references: EMBL:Z81055; PIDN:CAB02896.1; GSPDB:GN00022; CESP:F01G10.7
A;Experimental source: clone F01G10
C;Genetics:
A;Gene: CESP:F01G10.7
A;Map position: 4
A;Introns: 29/3; 64/1; 186/3; 347/2
C;Superfamily: triacylglycerol lipase, lingual


  Query Match            29.2%;  Score 655.5;  DB 2;  Length 426;
  Best Local Similarity   35.9%;  Pred. No. 6.6e-50;
  Matches  140;  Conservative   85;  Mismatches  144;  Indels   21;  Gaps    9;

Qy      20 LILVAYMFQRNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIP 79
           | |:   |  |:      : |:| | :| : || :| ||  | : | | | ||| ::|||
Db       6 LFLILLSFSINL------SLAIDDECYMTVPEIGKHFGYESEVHLVRTTDEYILELHRIP 59

Qy      80 -RGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWS 138
            :  :  ::  ||:| :||||:    :|| || | | ||: ||||||:|: ||||   |
Db      60 CKQNEKCDRSSKRPIVFMQHGLLADGFSWIPNLANQSAGFVFADAGFDIWIANSRGTPAS 119

Qy     139 RKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFST 198
           :||     :  :|| |:: :|: ||| | :: :|::| || :||:|:|||| : |   :
Db     120 QKHIGYGPENQKFWNFTWQQMSEFDLTASVDLVLKETKQEPLYYLGHSQGTMIMFSRLAE 179

Qy     199 MPELAQKIKMYFALAPIATVKHAKSP----GTKFLLLPDMMIKGLFGKKEF--LYQTRFL 252
           || |:||: : ||||:||| |        |||||   ::    | |:  :  |   | :
Db     180 NPEFAKKIRHFHALAPVATVSHIGGLFGLFGTKFLTYAEI----LLGRLPYSPLSIPRTV 235

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
           :::: |:| :  :  ||:  :  : | |    | ||  ||  || | |||::: || | |
Db     236 QKMISYMCSRFFMQNICTLDIGFIDG-NEKMFNQSRVGVYLCHTPAATSVKDLQHWIQLV 294

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVK-MLLSE 371
             |  :  ||:|::  |: :  ||||  | :  :  || ::    | |:: :|::  :||:
Db     295 KSQTVSKFDYGTD-GNIIEYGQPTPPEYDLTQINTPTYLYWSRDDILADTQDIRDSILSK 353

Qy     372 VTNLIYHK-NIPEWAHVDFIWGLDAPHRMY 400
           :   |     :| ::|:||::|  |   :|
Db     354 MNKTIAGSLELPHYSHMDFVFGTHAAFDLY 383


RESULT  10
JT0949
egg-specific protein - silkworm
C;Species: Bombyx mori (silkworm)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 26-Feb-1998
C;Accession: JT0949; A28527
R;Sato, Y.; Yamashita, O.
submitted to JIPID, September 1991
A;Reference number: JT0949
A;Accession: JT0949
A;Molecule type: DNA
A;Residues: 1-559 <SAT>
A;Experimental source: larva
A;Note: this protein is a homotrimer
R;Indrasith, L.S.; Sasaki, T.; Yamashita, O.
J. Biol. Chem. 263, 1045-1051, 1988
A;Title: A unique protease responsible for selective degradation of a yolk protein in Bo
A;Reference number: A28527; MUID:88087166
A;Accession: A28527
A;Molecule type: protein
A;Residues: 19-40;133-144,'T',146-152;229-248 <IND>
C;Genetics:
A;Map position: 19
C;Keywords: egg yolk; homotrimer
F;132-133/Cleavage site: Lys-Asn (egg 30.5K cysteine proteinase) #status experimental
F;228-229/Cleavage site: Arg-Asp (egg 30.5K cysteine proteinase) #status experimental


  Query Match            25.6%;  Score 574.5;  DB 2;  Length 559;
  Best Local Similarity   33.2%;  Pred. No. 1.4e-42;
  Matches  132;  Conservative   82;  Mismatches  152;  Indels   31;  Gaps    9;

Qy      32 NSVH---MPTKAVDPEAF-MNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKK 87
           |: |     |:  : | | :| :|:::   || ||: |||:||| |:| |||    | :
Db     171 NTFHDAISETQRENNEDFHLNATELLKKHQYPVEEHTVATDDGYHLTVLRIP-PTHQTRD 229

Qy      88 TGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSID 147
             :|| || |||:| | :|:   |: || ::|:|||:|||:|| ||| :|| | :
Db     230 DKKKPVALLMHGLLGSADDWLLMGPSKSLAYMLSDAGYDVWLGNVRGNKYSRSHVSKHPA 289

Qy     148 QDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIK 207
           ::|| || ||:|  ||||:|: :|  :|||:::|:|:||| |  |   |  |   :||
Db     290 LNDFWKFSNDEIALHDLPAIIDHVLDISGQERLHYIGHSQGATTFFALMSEQPSYNEKIV 349

Qy     208 MYFALAPIATVKHAKS-------PGTKFL-LLPDMMIKGLFGKKEFLYQTRFLRQLVIYL 259
             ||:||  : : :|        | :||   : | :  | |   : | :|
Db     350 SMHALSPIVYMNYVRSPLFRMIAPTSKFYQYIHDQVGHGAFEPGKHLIET---------F 400

Qy     260 CGQVILDQ-----ICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNS 314
           |    ::     :|:|:  :: | |  | :      |  ||  |||| : :  : | | |
Db     401 GGAACREKLGCRHVCNNLNYVISGINVYNQDADIVPVVMAHLPAGTSARVMKQYGQNVAS 460

Qy     315 GELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTN 374
           : | :::|:|| |::       | | :  :: |  ::     ||:|:|:|| |   : |
Db     461 HDFRKYNYGAET-NMKVYGASEPPSYDLSKVSAPVNLYHSHDAWLAHPKDVEKLQENLPN 519

Qy     375 LIYHKNIPEWAH---VDFIWGLDAPHRMYNEIIHLMQ 408
           :   :|| |  :|| :  ||  :| :::  ||
Db     520 VKQSFEVPEQQHFTDLDFQFSKKAPDTVYQKLMENMQ 556


RESULT  11
T39540
triglyceride lipase-cholesterol esterase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: T39540
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21862
A;Accession: T39540
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-443 <WOO>
A;Cross-references: EMBL:AL021748; PIDN:CAA16863.1; GSPDB:GN00067; SPDB:SPBC16A3.12c
A;Experimental source: strain 972h-; cosmid c16A3
C;Genetics:
A;Gene: SPDB:SPBC16A3.12c
A;Map position: 2
A;Introns: 11/2; 152/3; 282/1; 398/3
C;Superfamily: triacylglycerol lipase, lingual


  Query Match            22.5%;  Score 505;  DB 2;  Length 443;
  Best Local Similarity   29.6%;  Pred. No. 1.3e-36;
  Matches  125;  Conservative   91;  Mismatches  154;  Indels   52;  Gaps   13;

Qy       9 WIVSHRMEMW--LLILVAYMFQRNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVA 66
             :::  |:  |  : |:| | :  :  :    | ||  :   || || :  ||  ||: |
Db      35 YMIPERVREWFRVRIIVLYHYYISSKTTDGMTDAV--QKCRNIYEICEAFGYRVEEHLVR 92

Qy      67 TEDGYILSVNRIPRGLVQPKKT-GSRPVVLLQHGLVGGASNWIS-NLPNNSLGFILADAG 124
           |:| :|| ::||       ||::   | ||   |||:  :  |:: |    || |:| ::|
Db      93 TQDNFILCLHRI----THPKQSQHKREVVYCHHGLMTNSELWVAVNESERSLPFVLIESG 148

Qy     125 FDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVG 184
           :|||:||:||| :|||| |     :||| || |:|| ||:|  :::||::||:||: |:|
Db     149 YDVWLGNNRGNKYSRKHITYKPKDEEFWNFSLDDMAMFDIPDTVDYILRETGREKLNYIG 208

Qy     185 YSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLL--PDMMIKGLFGK 242
           :||||     | |  |:|  |: ::  |||    |   :    :::   | :|   |||:
```

```
Db    209 FSQGTAQAMAALSINPDLNDKVNIFIGLAPAYAPKGFSNYFVDYIVKVNPKIMYH-LFGR 267

Qy    243 KEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSV 302
          :  |    | :     :|   |  :|    : :|  :: :|:::::     ||  : :||
Db    268 RCLLPSVTFWQN----ICYPPIFVKIVDVSLKILFNWDLSNISLNQKLCGYAHLYSFSSV 323

Qy    303 QNILHWSQAVNSGELRAFD--------WGSE--------TKNLEKCNQPTPVRYRVRDMT 346
          ::::|| | : :   : :|         :||          | |: ||
Db    324 KSVVHWLQIIKNCTFQLYDDDMALLAGYGSRHYQVPLFPTNNI-KC-------------- 368

Qy    347 VPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHL 406
           |  :  ||:| | | | ::  |      :   :|  : |:||:|| |    ::  :|
Db    369 -PMLILWGGKDTLINMEVMRTALPPHAKEV---SIAHYEHLDFLWGQDVKEEVFPVVIDA 424

Qy    407 MQ 408
          ::
Db    425 LK 426


RESULT  12
T31611
hypothetical protein Y50E8A.g - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T31611
R;Steward, C.
submitted to the EMBL Data Library, September 1999
A;Reference number: Z21047
A;Accession: T31611
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1585 <WIL>
A;Cross-references: EMBL:AL117200; NID:e1549770; PIDN:CAB55050.1; CESP:Y50E8A.g
A;Experimental source: clone Y50E8A
C;Genetics:
A;Gene: CESP:Y50E8A.g
A;Introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2


  Query Match             21.3%;  Score 478.5;  DB 2;  Length 1585;
  Best Local Similarity   30.6%;  Pred. No. 1.8e-33;
  Matches  118;  Conservative   67;  Mismatches  172;  Indels   29;  Gaps    9;

Qy     44 EAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTG--SRPVVLLQHGLV 101
          |  :: :: | | ||  |:: | |:||| : : ||| |        |  ||||   |||
Db     20 EETLDAADTISHYGYTVEKHYVTTDDGYTVQLQRIPVGRDDRSILGCSKRPVVFFMHGLF 79

Qy    102 GGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEMAR 161
          | : ::: |||: |  :| |||||||||:|| ||  :   | : | :   || ||  | :
Db     80 GSSYHFLLNLPSQSAAYIFADAGFDVWLGNIRGTEYGLNHTSFSTNGVNFWNFSLYEHSH 139

Qy    162 FDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPEL----AQKIKMYFALAPIAT 217
          :||   | ::|  |  | ::|||:|||    : |: : |       ||:::||| | |
Db    140 YDLRQQIEYVLDYTRHESLFYVGHSQGTA---VMFARLAEADVTWQSKIRVFFALGPTAG 196

Qy    218 VKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRF----------LRQLVIYLCGQVILDQ 267
                |  | || :  :: |    :|    :|           :       |
Db    197 FMKPLMP---FTLLEENYLQALI---QFALDGKFGILPVEIPRAIASKFADFCSSKFFTF 250

Qy    268 ICSNIMLLLGGFNT-NNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSET 326
          :||    :  |  |   :| ||  :  :|  : ||  |:||| |     |||  | |:
Db    251 LCSAGFKVAAGIETLGQVNDSRIPIILSHFPSATSTLNLLHWMQIFKYHELRRLDLGT-A 309

Qy    327 KNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV--TNLIYHKNIPEW 384
          :||    |    |  : ::   | ::    | :::  ||: :: :     ||   ::  :
Db    310 RNLIAYGQKDAPRLEIGNIIAQTILYFSKDDQITDEVDVREIIMKQMGPGLIESYDLDHF 369

Qy    385 AHVDFIWGLDAPHRMYNEIIHLMQQE 410
           | ||| || |  :|  |:: : :|
Db    370 THFDFILGLRATDEVYKPIVYRIYKE 395


RESULT  13
T41053
triglyceride lipase-cholesterol esterase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T41053
R;Hilbert, H.; Duesterhoeft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, August 1998
A;Reference number: Z21967
A;Accession: T41053
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-467 <HIL>
A;Cross-references: EMBL:AL031324; PIDN:CAA20447.1; GSPDB:GN00068; SPDB:SPCC1672.09
A;Experimental source: strain 972h-; cosmid c1672
C;Genetics:
A;Gene: SPDB:SPCC1672.09
A;Map position: 3
A;Introns: 12/2; 164/3; 294/1


  Query Match             21.1%;  Score 475;  DB 2;  Length 467;
  Best Local Similarity   31.2%;  Pred. No. 6.4e-34;
  Matches  125;  Conservative   68;  Mismatches  158;  Indels   50;  Gaps   12;

Qy     39 KAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPR------GLVQPKKTGSRP 92
          ||:   |  :| |: :  ||  |:: | ||| ||| ::|| :      |  |||    |
Db     73 KAIHDAA--DIREMCKISGYYVEDHLVRTEDDYILCIHRISKDSPGRIGSPHPKKL---P 127

Qy     93 VVLLQHGLVGGASNWISNL-PNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEF 151
          ||   |||:  :  |: |: | | | | | : |:|||:||:||| :||:|        ||
Db    128 VVYCHHGLLMNSEVWVCNVDPRNCLVFDLVNKGYDVWLGNNRGNKYSRQHLRFDSTDKEF 187

Qy    152 WAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFA 211
          | || |: |::|:|  |::||: :|| |: |:|:|||   | : |  | |  ||    |
Db    188 WDFSIDDFAQYDIPDTIDYILKTSGQTKLTYIGFSQGTAQAFASLSIHPLLNDKINSLIA 247

Qy    212 LAPIATVKHAKSPGTKFLLLPDMMIKG-------LFGKKEFLYQTRFLRQLVIYLCGQVI 264
          |||      | ||      : |  :|      |||:|  |    | :  :
Db    248 LAP------AISPKGLHNRVVDAFVKARPSILFFLFGRKSILPSAGFWQSFL----APKF 297

Qy    265 LDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFD--- 321
          | : :  :  |  ::  |::  :  |  ||  : |||: ::|| | : | | | :|
Db    298 FDAVLAYCLSQLFNWSCQNISSYQRLVSFAHLYSYTSVKCLVHWFQIMRSAEFRMYDNDQ 357

Qy    322 WGSE--TKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHK 379
          | :   |  :    ||       ::  |  :  || | |    |:: :|: :   : |
Db    358 LGHDYFLKYYKAAKFPT------NNIRTPIYLIWGGSDSL---VDIQAMLNALPAEVEHV 408

Qy    380 NIPEWAHVDFIWG-------LDAPHRMYNEIIHLMQQEETN 413
          :  : |:| ||        :    |  :| |  : || :
Db    409 KVDSYEHLDMIWADTVKDYVIPPVLRRLRDIHHPPEHEEND 449


RESULT  14
E84526
probable lysosomal acid lipase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: E84526
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallor
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Vente
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: E84526
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-344 <STO>
A;Cross-references: GB:AE002093; NID:g4585908; PIDN:AAD25569.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g15230
A;Map position: 2
```

```
  Query Match           19.8%;  Score 445;  DB 2;  Length 344;
  Best Local Similarity  31.2%;  Pred. No. 1.9e-31;
  Matches  115;  Conservative   55;  Mismatches  124;  Indels   74;  Gaps   12;

Qy      50 SEIIQHQGYPCEEYEVATEDGYILSVNRI----PRGLVQPKKTGSRPVVLLQHGLVGGAS 105
           :::|    | | |: : |:|||||:: |:    ||         | | |||||||
Db      34 ADLIHPANYSCTEHSIQTKDGYILALQRVASLGPR-------LQSGPPVLLQHGLFMAGD 86

Qy     106 NWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLP 165
             |  | |  |||||||| |||||:|| ||  :|  | |||    ||| :|: ::| :||
Db      87 VWFLNSPKESLGFILADHGFDVWVGNVRGTRYSYGHVTLSDTDKEFWDWSWQDLAMYDLA 146

Qy     166 AVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPG 225
           :| : |      ||: ||:|||| | | |  | | :|: ::    | ||: : |  :|
Db     147 EMIQY-LYSISNSKIFLVGHSQGTIMSFAAL-TQPHVAEMVEAAALLCPISYLDHVTAP- 203

Qy     226 TKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGFNTNNMN 285
               | : |:        |::          ||||
Db     204 -----LVERMV--------FMH----------------LDQI---------------- 216

Qy     286 MSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDM 345
               |  :    :||:|| |  | : |    :|:|   |||       |  : :  :
Db     217 ----EYYLDYEPHPSSVKNIRHLFQMIRKGTFAQYDYG-YFKNLRTYGLSKPPEFILSHI 271

Qy     346 TVPTAMWT--GGQDWLSNPEDVKMLLSEVTN---LIYHKNIPEWAHVDFIWGLDAPHRMY 400
              ||   || | |::  ||:  |:|: :   |:|   : :: |:||: |  |   :|
Db     272 PASLPMWMGYGGTDGLADVTDVEHTLAELPSSPELLY---LEDYGHIDFVLGSSAKEDVY 328

Qy     401 NEIIHLMQ 408
             :|   :
Db     329 KHMIQFFR 336


RESULT  15
S37969
probable triacylglycerol lipase (EC 3.1.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKL140w
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 06-Feb-1998
C;Accession: S37969; S39000; S22278
R;Ramezani Rad, M.; Xu, G.; Kirchrath, L.; Fritz, C.; Keuchel, H.; Hollenberg, C.P.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37953
A;Accession: S37969
A;Molecule type: DNA
A;Residues: 1-548 <RAM>
A;Cross-references: EMBL:Z28140; NID:g486238; PID:g486239; MIPS:YKL140w
A;Experimental source: strain S288C
R;Abraham, P.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S38996
A;Accession: S39000
A;Molecule type: DNA
A;Residues: 1-548 <ABR>
A;Cross-references: EMBL:Z25464; NID:g396434; PID:g396439
R;Abraham, P.R.; Mulder, A.; van't Riet, J.; Planta, R.J.; Raue, H.A.
Yeast 8, 227-238, 1992
A;Title: Molecular cloning and physical analysis of an 8.2 kb segment of chromosome XI c
A;Reference number: S22274; MUID:92245761
A;Accession: S22278
A;Molecule type: DNA
A;Residues: 1-398 <AB2>
A;Cross-references: GB:S97962
C;Genetics:
A;Gene: SGD:TGL1
A;Cross-references: SGD:S0001623; MIPS:YKL140w
A;Map position: 11L
C;Keywords: carboxylic ester hydrolase

  Query Match           18.8%;  Score 422.5;  DB 2;  Length 548;
  Best Local Similarity  31.6%;  Pred. No. 3.5e-29;
  Matches  119;  Conservative   66;  Mismatches  127;  Indels   65;  Gaps   15;

Qy      61 EEYEVATEDGYILSVNRIPRGLVQP--KKTGSRPVVLLQHGLVGGASNWISNLP-NNSLG 117
           |:: | ||| |||:::|||     |  |   :  || | |||:  :  |  |:  : :|
Db      78 EDHLVRTEDNYILTLHRIP-----PISKNRFNNKVVYLHHGLLMCSDVWCCNIERHKNLP 132

Qy     118 FILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQ 177
           |:| | |:||||||:||| :|  |       ::|| || || | ||:|  | |||  |
Db     133 FVLHDLGYDVWMGNNRGNKYSTAHLNKPPKSNKFWDFSIDEFAFFDIPNSIEFILDITKV 192

Qy     178 EKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVK--HAK--------SPGTK 227
           :|:  :|:|||:   | |||   :| :|:  : |:||  | |  | :        |||
Db     193 DKVICIGFSQGSAQMFAAFSLSEKLNRKVSHFIAIAPAMTPKGLHNRIVDTLAKSSPGFM 252

Qy     228 FLLLPDMMIKGLFGKKEFLYQT----RFLRQLVIYLCGQVILDQICSNIMLLLGGFNTNN 283
           :|        ||:|  |       | |   :  ||    :| | : |: .    ||  :
Db     253 YL---------FFGRKIVLPSAVIWQRTLHPTLFNLC----ID-IANKIL-----FNWKS 293

Qy     284 MNM-------SRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFD-----WGSETKNLEK 331
           |:        | | :|:      |||::|:|| | : | : : |:       | |:  :
Db     294 FNILPRQKIASYAKLYST-----TSVKSIVHWFQILRSQKFQMFEESDNMLNSLTRPYQI 348

Qy     332 CNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIW 391
           | ||       :: :|  :  || | | : : :|  |    | ::   :  : |:| ||
Db     349 ANFPTRT-----NIKIPILLIYGGIDSLVDIDVMKKNLP--FNSVFDVKVDNYEHLDLIW 401

Qy     392 GLDAPHRMYNEIIHLMQ 408
           | ||   :  :::  ::
Db     402 GKDADTLVIAKVLRFIE 418


Search completed: January 30, 2002, 23:02:30
Job time: 2911 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 30, 2002, 23:01:24 ; Search time 50.02 Seconds
(without alignments)
310.060 Million cell updates/sec

Title: US-09-578-063-47
Perfect score: 2247
Sequence: 1 MLETLSRQWIVSHRMEMWLL.....IHLMQQEETNLSQGRCEAVL 423

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1289	57.4	399	1	LICH_HUMAN	P38571 homo sapien
2	1164.5	51.8	395	1	LIPG_RAT	P04634 rattus norv
3	1154	51.4	398	1	LIPG_CANFA	P80035 canis famil
4	1139	50.7	397	1	LIPG_BOVIN	Q29458 bos taurus
5	1119	49.8	398	1	LIPG_HUMAN	P07098 homo sapien
6	663	29.5	394	1	LIP3_DROME	O46108 drosophila
7	546	24.3	439	1	LIP1_DROME	O46107 drosophila
8	422.5	18.8	548	1	TGL1_YEAST	P34163 saccharomyc
9	126	5.6	987	1	YD94_METJA	Q58789 methanococc
10	105	4.7	430	1	ORC2_CAEEL	Q21037 caenorhabdi
11	101	4.5	436	1	PAFA_CAVPO	P70683 c platelet-
12	96.5	4.3	287	1	Y193_HAEIN	Q57427 haemophilus
13	95.5	4.3	1259	1	LIN1_HUMAN	P08547 homo sapien
14	93.5	4.2	278	1	PRXC_STRAU	O31168 streptomyce
15	93	4.1	569	1	PYRD_PLAFA	Q08210 plasmodium
16	92	4.1	748	1	PT1P_ECOLI	P37177 escherichia
17	91	4.0	370	1	P2C2_SCHPO	Q09172 schizosacch
18	90.5	4.0	1037	1	ACRD_ECOLI	P24177 escherichia
19	90	4.0	277	1	BPA2_STRAU	P29715 streptomyce
20	90	4.0	1122	1	YG3C_YEAST	P53280 saccharomyc
21	90	4.0	1136	1	POLG_OMV	P20234 o genome po
22	89	4.0	330	1	EXOA_RHIME	P33691 rhizobium m
23	88.5	3.9	185	1	YC2H_BACSU	O31482 bacillus su
24	88.5	3.9	869	1	CFAC_ECOLI	P25733 escherichia
25	87	3.9	473	1	SCRB_LACLA	Q04937 lactococcus
26	86.5	3.8	382	1	YDHH_HAEIN	P44861 haemophilus
27	86.5	3.8	665	1	INVA_SALTY	P35657 salmonella
28	86.5	3.8	1882	1	Y468_MYCPN	P75109 mycoplasma
29	86	3.8	525	1	ACU8_NEUCR	P15937 neurospora
30	86	3.8	1866	1	VGNB_CPMV	P03600 cowpea mosa
31	85.5	3.8	428	1	YE63_SCHPO	O14249 schizosacch
32	85	3.8	429	1	ARB2_ECOLI	P52146 escherichia
33	85	3.8	837	1	UBPA_DICDI	P54201 dictyosteli
34	85	3.8	4128	1	PRKD_MOUSE	P97313 mus musculu
35	84.5	3.8	875	1	AMD2_XENLA	P12890 xenopus lae
36	84	3.7	451	1	YB27_YEAST	P38295 saccharomyc
37	84	3.7	804	1	SYL_BACSU	P36430 bacillus su
38	83.5	3.7	350	1	E13B_PRUPE	P52408 prunus pers
39	83.5	3.7	375	1	NDR3_MOUSE	Q9qyf9 mus musculu
40	83.5	3.7	389	1	AAT_PYRAB	Q9v012 pyrococcus
41	83.5	3.7	456	1	YP95_YEAST	Q02891 saccharomyc
42	83.5	3.7	520	1	C11A_BOVIN	P00189 bos taurus
43	83.5	3.7	520	1	C11A_PIG	P10612 sus scrofa
44	83.5	3.7	690	1	VGLH_HSV7J	P52353 human herpe
45	83	3.7	266	1	THID_SALTY	P55882 salmonella

ALIGNMENTS

```
RESULT   1
LICH_HUMAN
ID   LICH_HUMAN     STANDARD;      PRT;   399 AA.
AC   P38571;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   LYSOSOMAL ACID LIPASE/CHOLESTERYL ESTER HYDROLASE PRECURSOR
DE   (EC 3.1.1.13) (LAL) (ACID CHOLESTERYL ESTER HYDROLASE) (STEROL
DE   ESTERASE) (LIPASE A) (CHOLESTERYL ESTERASE).
GN   LIPA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 196-212; 277-297 AND 305-315.
RX   MEDLINE=92042192; PubMed=1718995;
RA   Anderson R.A., Sando G.N.;
RT   "Cloning and expression of cDNA encoding human lysosomal acid
RT   lipase/cholesteryl ester hydrolase. Similarities to gastric and
RT   lingual lipases.";
RL   J. Biol. Chem. 266:22479-22484(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=94155897; PubMed=8112342;
RA   Ameis D., Merkel M., Eckerskorn C., Greten H.;
RT   "Purification, characterization and molecular cloning of human
RT   hepatic lysosomal acid lipase.";
RL   Eur. J. Biochem. 219:905-914(1994).
RN   [3]
RP   VARIANT CESD/WD PRO-200.
RX   MEDLINE=94195814; PubMed=8146180;
RA   Anderson R.A., Byrum R.S., Coates P.M., Sando G.N.;
RT   "Mutations at the lysosomal acid cholesteryl ester hydrolase gene
RT   locus in Wolman disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:2718-2722(1994).
RN   [4]
RP   VARIANTS CESD ARG-129 AND PRO-129.
RX   MEDLINE=98295576; PubMed=9633819;
RA   Ries S., Buechler C., Schindler G., Aslanidis C., Ameis D., Gasche C.,
RA   Jung N., Schambach A., Fehringer P., Vanier M.T., Belli D.C.,
RA   Greten H., Schmitz G.;
RT   "Different missense mutations in histidine-108 of lysosomal acid
RT   lipase cause cholesteryl ester storage disease in unrelated compound
RT   heterozygous and hemizygous individuals.";
RL   Hum. Mutat. 12:44-51(1998).
CC   -!- FUNCTION: CRUCIAL FOR THE INTRACELLULAR HYDROLYSIS OF CHOLESTERYL
CC       ESTERS AND TRIGLYCERIDES THAT HAVE BEEN INTERNALIZED VIA RECEPTOR-
CC       MEDIATED ENDOCYTOSIS OF LIPOPROTEIN PARTICLES. IMPORTANT IN
CC       MEDIATING THE EFFECT OF LDL (LOW DENSITY LIPOPROTEIN) UPTAKE ON
CC       SUPPRESSION OF HYDROXYMETHYLGLUTARYL-COA REDUCTASE AND ACTIVATION
CC       OF ENDOGENOUS CELLULAR CHOLESTERYL ESTER FORMATION.
CC   -!- CATALYTIC ACTIVITY: A STERYL ESTER + H(2)O = A STEROL + A FATTY
CC       ACID.
```

```
CC   -!- SUBCELLULAR LOCATION: LYSOSOMAL.
CC   -!- DISEASE: DEFECTS IN LIPA ARE THE CAUSE OF THE SEVERE INFANTILE-
CC       ONSET WOLMAN DISEASE (WD) AND THE MILDER LATE-ONSET CHOLESTERYL
CC       ESTER STORAGE DISEASE (CESD).
CC   -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC       HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M74775; AAA59519.1; -.
DR   EMBL; U04285; AAB60327.1; -.
DR   EMBL; U04286; AAB60327.1; JOINED.
DR   EMBL; U04287; AAB60327.1; JOINED.
DR   EMBL; U04288; AAB60327.1; JOINED.
DR   EMBL; U04290; AAB60327.1; JOINED.
DR   EMBL; U04291; AAB60327.1; JOINED.
DR   EMBL; U04292; AAB60327.1; JOINED.
DR   EMBL; U04293; AAB60327.1; JOINED.
DR   EMBL; X76488; CAA54026.1; -.
DR   PIR; A39315; A39315.
DR   MIM; 278000; -.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
KW   Hydrolase; Lipid degradation; Glycoprotein; Signal; Lysosome;
KW   Disease mutation; Polymorphism.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    399       LYSOSOMAL ACID LIPASE/CHOLESTERYL ESTER
FT                                HYDROLASE.
FT   ACT_SITE    174    174       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    374    374       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     36     36       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     72     72       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    101    101       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    273    273       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    321    321       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT      16     16       P -> T.
FT                                /FTId=VAR_004247.
FT   VARIANT     129    129       H -> P (IN CESD).
FT                                /FTId=VAR_004248.
FT   VARIANT     129    129       H -> R (IN CESD).
FT                                /FTId=VAR_004249.
FT   VARIANT     200    200       L -> P (IN CESD AND WD).
FT                                /FTId=VAR_004250.
SQ   SEQUENCE   399 AA;  45415 MW;  55F27391306B609A CRC64;


  Query Match             57.4%;  Score 1289;  DB 1;  Length 399;
  Best Local Similarity   63.5%;  Pred. No. 3.8e-104;
  Matches  235;  Conservative   55;  Mismatches   80;  Indels    0;  Gaps    0;

Qy      40 AVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLLQHG 99
           |||||  ||:|||| : |:| ||| | ||||||| :|||| |       | :||| ||||
Db      28 AVDPETNMNVSEIISYWGFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHG 87

Qy     100 LVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEM 159
           |:  :|||::|| |:||||||||||||||||||||| |||||||||: ||||||||||||
Db      88 LLADSSNWVTNLANSSLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEM 147

Qy     160 ARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVK 219
           |::|||| ||||| ||||||::||||:|||||:|||||| :||||::|||:||| |:|:|
Db     148 AKYDLPASINFILNKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVA 207

Qy     220 HAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGF 279
```

```
           || |   ||| :|| ||| |||| |: ||: |  ::|  ||| ::| |:  || ||
Db     208 FCTSPMAKLGRLPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGF 267

Qy     280 NTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVR 339
           |  |:|||  ||  |: |||||||:|||||||   : :|||||  ||    || |
Db     268 NERNLNMSRVDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPT 327

Qy     340 YRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHRM 399
           | |:|| ||||:|:|| |||::  || :||:::|||::|::|||| |:|||||||| |:
Db     328 YNVKDMLVPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRL 387

Qy     400 YNEIIHLMQQ 409
           ||:||:||::
Db     388 YNKIINLMRK 397


RESULT   2
LIPG_RAT
ID   LIPG_RAT       STANDARD;      PRT;   395 AA.
AC   P04634;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence.update)
DT   15-JUL-1999 (Rel. 38, Last annotatiqn update)
DE   TRIACYLGLYCEROL LIPASE, LINGUAL PRECURSOR (EC 3.1.1.3) (LINGUAL
DE   LIPASE).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85215587; PubMed=3839077;
RA   Docherty A.J.P., Bodmer M.W., Angal S., Verger R., Riviere C.,
RA   Lowe P.A., Lyons A., Emtage J.S., Harris T.J.R.;
RT   "Molecular cloning and nucleotide sequence of rat lingual lipase
RT   cDNA.";
RL   Nucleic Acids Res. 13:1891-1903(1985).
CC   -!- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +
CC       A FATTY ACID ANION.
CC   -!- MISCELLANEOUS: LINGUAL LIPASE IS SECRETED BY THE SEROUS (VON
CC       EBNER'S) GLANDS AT THE BACK OF THE RAT TONGUE.
CC   -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC       HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02309; CAA26179.1; -.
DR   EMBL; A01157; CAA00136.1; -.
DR   PIR; A23045; LIRTT.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
KW   Hydrolase; Lipid degradation; Glycoprotein; Signal.
FT   SIGNAL        1     18
FT   CHAIN        19    395       TRIACYLGLYCEROL LIPASE, LINGUAL.
FT   ACT_SITE    171    171       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    371    371       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     33     33       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    270    270       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   395 AA;  44588 MW;  E601854A923522EA CRC64;


  Query Match             51.8%;  Score 1164.5;  DB 1;  Length 395;
```

```
  Best Local Similarity   54.9%;  Pred. No. 2.3e-93;
  Matches  217;  Conservative   69;  Mismatches  108;  Indels    1;  Gaps    1;

Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||||:: : :      : :       :||| ||||::| : ||||:|||| |||||||
Db       1 MWLLLITSVISTFGGAHGLFGKLGPGNPEANMNISQMITYWGYPCQEYEVVTEDGYILGV 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |    :  | |||| |||||:  |:|||:|||||| |:|||||:|||:||||||
Db      61 YRIPHGKNNSENIGKRPVVYLQHGLIASATNWIANLPNNSLAFMLADAGYDVWLGNSRGN 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           ||||:   | |  ||||||:||||::|||| ||||:||||||||:|||:|||||:||||
Db     121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATINFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQL 255
           ||| | ||:||| ::|||||:|||||: :||  |   :|  : | :|||| ||  | |   |
Db     181 FSTNPTLAKKIKTFYALAPVATVKYTQSPLKKISFIPTFLFKLMFGKKMFLPHTYFDDFL 240

Qy     256 VIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSG 315
             :| : :|| :||| : :  ||:  |:|:||  ||  |  |||||: |||:| | ||
Db     241 GTEVCSREVLDLLCSNTLFIFCGFDKKNLNVSRFDVYLGHNPAGTSVQDFLHWAQLVRSG 300

Qy     316 ELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNL 375
           : :||:||| ::|:   || ||  | |  |||| |:| || | |::|:|| ||| :::||
Db     301 KFQAFNWGSPSQNMLHYNQKTPPEYDVSAMTVPVAVWNGGNDILADPQDVAMLLPKLSNL 360

Qy     376 IYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ::|| |  : |:|||| :|||  :|||:| :| ::
Db     361 LFHKEILAYNHLDFIWAMDAPQEVYNEMISMMAED 395


RESULT   3
LIPG_CANFA
ID   LIPG_CANFA     STANDARD;      PRT;   398 AA.
AC   P80035; O02857;
DT   01-NOV-1991 (Rel. 20, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   TRIACYLGLYCEROL LIPASE, GASTRIC PRECURSOR (EC 3.1.1.3) (GASTRIC
DE   LIPASE) (GL).
GN   LIPF.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Stomach;
RA   Vaganay S., Joliff G., Bertaux O., Toselli E., Devignes M.D.,
RA   Benicourt C.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 20-59.
RX   MEDLINE=92037652; PubMed=1935982;
RA   Carriere F., Moreau H., Raphel V., Laugier R., Benicourt C.,
RA   Junien J.-L., Verger R.;
RT   "Purification and biochemical characterization of dog gastric
RT   lipase.";
RL   Eur. J. Biochem. 202:75-83(1991).
CC   -!- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +
CC       A FATTY ACID ANION.
CC   -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC       HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y13899; CAA74198.1; -.
DR   PIR; S19539; S19539.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
KW   Hydrolase; Lipid degradation; Glycoprotein; Signal.
FT   SIGNAL       1     19
FT   CHAIN       20    398       TRIACYLGLYCEROL LIPASE, GASTRIC.
FT   ACT_SITE    172   172       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    372   372       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     34    34       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     99    99       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    271   271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    327   327       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     39    39       I -> T (IN REF. 2).
SQ   SEQUENCE   398 AA;  45130 MW;  E04D62F7518E386C CRC64;


  Query Match             51.4%;  Score 1154;  DB 1;  Length 398;
  Best Local Similarity   54.3%;  Pred. No. 1.9e-92;
  Matches  216;  Conservative   71;  Mismatches   99;  Indels   12;  Gaps    5;

Qy      17 MWLLILVAYMFQRNVNSVH------MPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDG 70
           ||||:  | :    : : |       ||   :||  ||||::| : ||| ||||| ||||
Db       1 MWLLLTAASVIS-TLGTTHGLFGKLHPT---NPEVTMNISQMITYWGYPAEEYEVVTEDG 56

Qy      71 YILSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMG 130
           ||| ::||| |    :  | |||  ||||:  |:|||||||||| |||||||:|||:|
Db      57 YILGIDRIPYGRKNSENIGRRPVAFLQHGLLASATNWISNLPNNSLAFILADAGYDVWLG 116

Qy     131 NSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTT 190
           ||||| |:|::   | |  ||||||:||||::|||| |:|||:||||:|::|||:|||||
Db     117 NSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFILKKTGQDKLHYVGHSQGTT 176

Qy     191 MGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTR 250
           :|||||||  |:||::|| ::||||:||||: ::   | :|:|  : | :|| | | |
Db     177 IGFIAFSTNPKLAKRIKTFYALAPVATVKYTETLLNKLMLVPSFLFKLIFGNKIF-YPHH 235

Qy     251 FLRQ-LVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWS 309
           |  | |   :| :  :| :|||  : :: ||:| |:|||  || :|  ||||||:||||
Db     236 FFDQFLATEVCSRETVDLLCSNALFIICGFDTMNLNMSRLDVYLSHNPAGTSVQNVLHWS 295

Qy     310 QAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLL 369
           ||| ||: :|||||  :|:   :|  |  | : || || |:| || | |::| || :||
Db     296 QAVKSGKFQAFDWGSPVQNMMHYHQSMPPYYNLTDMHVPIAVWNGGNDLLADPHDVDLLL 355

Qy     370 SEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLM 407
           |:: |||||: || : |:|||| :||| .:||||: :|
Db     356 SKLPNLIYHRKIPPYNHLDFIWAMDAPQAVYNEIVSMM 393


RESULT   4
LIPG_BOVIN
ID   LIPG_BOVIN     STANDARD;      PRT;   397 AA.
AC   Q29458;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   TRIACYLGLYCEROL LIPASE, PREGASTRIC PRECURSOR (EC 3.1.1.3) (PREGASTRIC
DE   LIPASE) (GL) (PREGASTRIC ESTERASE) (PGE).
GN   LIPF.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Tongue serous gland;
RX   MEDLINE=95011625; PubMed=7926811;
RA   Timmermans M.Y.J., Kupers L.P., Teuchy H.;
```

```
RT   "The cDNA sequence encoding bovine pregastric esterase.";
RL   Gene 147:259-262(1994).
CC   -!- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +
CC       A FATTY ACID ANION.
CC   -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC       HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L26319; AAA57037.1; -.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
KW   Hydrolase; Lipid degradation; Glycoprotein; Signal.
FT   SIGNAL       1     19       BY SIMILARITY.
FT   CHAIN       20    397       TRIACYLGLYCEROL LIPASE, PREGASTRIC.
FT   ACT_SITE   171    171       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   371    371       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    33     33       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   270    270       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   326    326       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   397 AA;  45231 MW;  F68977DED585EE36 CRC64;
```

```
  Query Match           50.7%;  Score 1139;  DB 1;  Length 397;
  Best Local Similarity 53.7%;  Pred. No. 3.7e-91;
  Matches  212;  Conservative   68;  Mismatches  113;  Indels    2;  Gaps    2;

Qy      18 WLLILVAYMFQRNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNR 77
           |||: | ::          :   | :||| ||:|::| : ||| | ::| | ||||| | |
Db       3 WLLVTVCFIHMSGNAFCFLGKIAKNPEASMNVSQMISYWGYPSEMHKVITADGYILQVYR 62

Qy      78 IPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAW 137
           || |     | |||| |||||:| |:|||||| |||||:|||||:|||:|||||| |
Db      63 IPHGKNNANHLGQRPVVFLQHGLLGSATNWISNLPKNSLGFLLADAGYDVWLGNSRGNTW 122

Qy     138 SRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFS 197
           :::|   | |  ||||||:|||| :|||: |:|||::|||:|::|||:|||||:||||||
Db     123 AQEHLYYSPDSPEFWAFSFDEMAEYDLPSTIDFILRRTGQKKLHYVGHSQGTTIGFIAFS 182

Qy     198 TMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LV 256
           | | ||:|||:::||||:||||: ||   |  |:|  : | :|| | | |   || | |
Db     183 TSPTLAEKIKVFYALAPVATVKYTKSLFNKLALIPHFLFKIIFGDKMF-YPHTFLEQFLG 241

Qy     257 IYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGE 316
           : :| :  || :| | :  : | :  | ||||  || ||  ||||||| ||| ||| ||:
Db     242 VEMCSRETLDVLCKNALFAITGVDNKNFNMSRLDVYIAHNPAGTSVQNTLHWRQAVKSGK 301

Qy     317 LRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLI 376
           :|||||:  :||   :|||| | :  | || |:|:   | |::|:||  |||:::|||
Db     302 FQAFDWGAPYQNLMHYHQPTPPIYNLTAMNVPIAVWSADNDLLADPQDVDFLLSKLSNLI 361

Qy     377 YHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQEE 411
           ||| || : |:|||| :|||  :||||: || ::
Db     362 YHKEIPNYNHLDFIWAMDAPQEVYNEIVSLMAEDK 396
```

```
RESULT   5
LIPG_HUMAN
ID   LIPG_HUMAN     STANDARD;      PRT;   398 AA.
AC   P07098;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   TRIACYLGLYCEROL LIPASE, GASTRIC PRECURSOR (EC 3.1.1.3) (GASTRIC
DE   LIPASE) (GL).
GN   LIPF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87299724; PubMed=3304425;
RA   Bodmer M.W., Angal S., Yarranton G.T., Harris T.J.R., Lyons A.,
RA   King D.J., Pieroni G., Riviere C., Verger R., Lowe P.A.;
RT   "Molecular cloning of a human gastric lipase and expression of the
RT   enzyme in yeast.";
RL   Biochim. Biophys. Acta 909:237-244(1987).
RN   [2]
RP   SEQUENCE OF 20-45.
RX   MEDLINE=89325292; PubMed=2753032;
RA   Bernbaeck S., Blaeckberg L.;
RT   "Human gastric lipase. The N-terminal tetrapeptide is essential for
RT   lipid binding and lipase activity.";
RL   Eur. J. Biochem. 182:495-499(1989).
CC   -!- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +
CC       A FATTY ACID ANION.
CC   -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC       HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X05997; CAA29413.1; -.
DR   EMBL; X05997; CAA29414.1; ALT_INIT.
DR   EMBL; A01046; CAA00125.1; -.
DR   EMBL; A12714; CAA01053.1; -.
DR   PIR; S04942; S04942.
DR   PIR; S07145; S07145.
DR   MIM; 601980; -.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
KW   Hydrolase; Lipid degradation; Glycoprotein; Signal.
FT   SIGNAL       1     19
FT   CHAIN       20    398       TRIACYLGLYCEROL LIPASE, GASTRIC.
FT   ACT_SITE   172    172       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   372    372       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    34     34       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    99     99       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   271    271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   327    327       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   398 AA;  45237 MW;  CD3EE1621C014F0F CRC64;
```

```
  Query Match           49.8%;  Score 1119;  DB 1;  Length 398;
  Best Local Similarity 52.2%;  Pred. No. 2e-89;
  Matches  210;  Conservative   73;  Mismatches  107;  Indels   12;  Gaps    2;

Qy      17 MWLLILVAYMFQRNVNSVHMPTKAV-------DPEAFMNISEIIQHQGYPCEEYEVATED 69
           ||||: :| :       ||   |  :        ||  ||||::| : ||| ||||| |||
Db       1 MWLLLTMASLI-----SVLGTTHGLFGKLHPGSPEVTMNISQMITYWGYPNEEYEVVTED 55

Qy      70 GYILSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWM 129
           |||| ||||| |       || |||| |||||:  |:|||||||||| |||||||:|||:
Db      56 GYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSLAFILADAGYDVWL 115

Qy     130 GNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGT 189
           |||||| |:|::   | |  ||||||:||||::|||| |:||::|||||:::|||:||||
Db     116 GNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKKTGQKQLHYVGHSQGT 175
```

```
Qy     190 TMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQT 249
           |:|||||| | ||::|| ::||||:||||: ||   |   :|  : | :|| | |
Db     176 TIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVPQSLFKFIFGDKIFYPHN 235

Qy     250 RFLRQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWS 309
             | : |   :| : :|: :||| : :: ||:: | | ||  || :|  |||||||: ||:
Db     236 FFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLDVYLSHNPAGTSVQNMFHWT 295

Qy     310 QAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLL 369
           ||| ||: :|:|||| :|   :|  | | | || |:| ||:| |::|:|| :||
Db     296 QAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVPIAVWNGGKDLLADPQDVGLLL 355

Qy     370 SEVTNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQEE 411
             :: |||||| || : |:|||| :|||  :||:|: :: :::
Db     356 PKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSMISEDK 397
```

```
RESULT   6
LIP3_DROME
ID   LIP3_DROME     STANDARD;      PRT;   394 AA.
AC   O46108;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   LIPASE 3 PRECURSOR (EC 3.1.1.-).
GN   LIP3 OR CG8823.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=98227315; PubMed=9566193;
RA   Pistillo D., Manzi A., Tino A., Pilo Boyl P., Graziani F., Malva C.;
RT   "The Drosophila melanogaster lipase homologs: a gene family with
RT   tissue and developmental specific expression.";
RL   J. Mol. Biol. 276:877-885(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- TISSUE SPECIFICITY: FAT BODY.
CC   -!- DEVELOPMENTAL STAGE: ONLY AT LARVAL STAGES.
CC   -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC       HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.). ALSO SIMILAR TO
CC       LEPIDOPTERAN EGG-SPECIFIC AND YOLK PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y14367; CAA74737.1; -.
DR   EMBL; AE003699; AAF54935.1; -.
DR   FlyBase; FBgn0023495; Lip3.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
KW   Hydrolase; Lipid degradation; Signal; Glycoprotein.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21    394       LIPASE 3.
FT   ACT_SITE   164    164       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   369    369       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD   131    131       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   394 AA;  44901 MW;  A718D1D743673802 CRC64;
```

```
  Query Match           29.5%;  Score 663;  DB 1;  Length 394;
  Best Local Similarity 37.0%;  Pred. No. 6.3e-50;
  Matches  137;  Conservative   80;  Mismatches  139;  Indels   14;  Gaps    7;

Qy      51 EIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPK--KTGSRPVVLLQHGLVGGASNWI 108
           | |:  ||| | :|| | | |||:::|||     ||  :: :|||  | ||::  :|:|:
Db      30 ERIEDDGYPMERHEVVTSDNYILTMHRIP---YSPKTGESSNRPVAFLMHGMLSSSSDWV 86

Qy     109 SNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVI 168
             | || ::|||||:||||||:||| :|: ||          || ||::|:  :|:||:|
Db      87 LMGPERSLAYMLADAGYDVWMGNARGNTYSKAHKYWPTYWQIFWNFSWNEIGMYDVPAMI 146

Qy     169 NFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTK- 227
           :::| ||||::: |||:|||||:  :  |  ||   |||    | | | : : ||| |:
Db     147 DYVLAKTGQQQVQYVGHSQGTTVYLVMVSERPEYNDKIKSAHLLGPAAYMGNMKSPLTRA 206

Qy     228 ---FLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVI-LDQICSNIMLLLGGFNTNN 283
              |  |: ::: : |  ||:   :| : | | :|       :|:| : |:|||:|
Db     207 FAPILGQPNAIVE-VCGSMEFMPSNKFKQDLGIEMCQATSPYADMCANEIFLIGGYDTEQ 265

Qy     284 MNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVR 343
           ::       | : || ||   ||: |  |||: | ||: :  :|  :     |  |:::
Db     266 LDYELLEHIKATSPAGASVNQNLHFCQEYNSGKFRKFDY-TALRNPYEYGSYFPPDYKLK 324

Qy     344 DMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIP--EWAHVDFIWGLDAPHRMYN 401
           .:   |  :: |  ||: :  ||: |  |: |:     :|  :|||:|||| :|   :|:
Db     325 NAKAPVLLYYGANDWMCDVSDVRKLRDELPNMALDYLVPFEKWAHLDFIWGTEARKYVYD 384

Qy     402 EIIHLMQQEE 411
           |::  ||  |
Db     385 EVLKQMQSYE 394
```

```
RESULT   7
LIP1_DROME
ID   LIP1_DROME     STANDARD;      PRT;   439 AA.
AC   O46107; Q9VKR6;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   LIPASE 1 PRECURSOR (EC 3.1.1.-).
GN   LIP1 OR CG7279.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE OF 7-433 FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=98227315; PubMed=9566193;
RA   Pistillo D., Manzi A., Tino A., Pilo Boyl P., Graziani F., Malva C.;
RT   "The Drosophila melanogaster lipase homologs: a gene family with
RT   tissue and developmental specific expression.";
RL   J. Mol. Biol. 276:877-885(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: COULD BE A DIGESTIVE ENZYME.
CC   -!- TISSUE SPECIFICITY: OVARIES.
CC   -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC       HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.). ALSO SIMILAR TO
CC       LEPIDOPTERAN EGG-SPECIFIC AND YOLK PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y14366; CAA74736.1; -.
DR   EMBL; AE003629; AAF52994.1; -.
DR   FlyBase; FBgn0024747; Lip1.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; FALSE_NEG.
KW   Hydrolase; Lipid degradation; Signal; Glycoprotein.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25    439       LIPASE 1.
FT   DOMAIN      30     44       POLY-GLU.
FT   ACT_SITE   197    197       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   393    393       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD   124    124       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   346    346       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   379    379       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   426    426       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    10     10       L -> I (IN REF. 1).
FT   CONFLICT   213    213       Y -> F (IN REF. 1).
FT   CONFLICT   412    412       Q -> E (IN REF. 1).
SQ   SEQUENCE   439 AA;  50660 MW;  9E32E20BEAE93E3F CRC64;


  Query Match             24.3%;  Score 546;  DB 1;  Length 439;
  Best Local Similarity   31.8%;  Pred. No. 9.9e-40;
  Matches  123;  Conservative   83;  Mismatches  157;  Indels   24;  Gaps    7;

Qy      39 KAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPVVLLQH 98
           |  :  :: ::: ::|   ||  | : | |||||||:::||       :| |: |  ||||
Db      58 KNIKQDSTLSVDKLIAKYGYESEVHHVTTEDGYILTMHRI-------RKQGA-PPFLLQH 109

Qy      99 GLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDE 158
           |||  :: ::   || || ::|||  :|||:||:||| :|| | ||  |: :|| ||: |
Db     110 GLVDSSAGFVVMGPNVSLAYLLADHNYDVWLGNARGNRYSRNHTTLDPDESKFWDFSWHE 169

Qy     159 MARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATV 218
           :  :||||:|: :|: ||  |::| |:||| |  |:  |  |    |:  ||||
Db     170 IGMYDLPAMIDHVLKVTGFPKLHYAGHSQGCTSFFVMCSMRPAYNDKVVSMQALAPAVYA 229

Qy     219 KHAKS-PGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLG 277
           |  :  |  : : |    : |   :: |  : ||| ::          :::|   :  :
Db     230 KETEDHPYIRAISLYFNSLVGSSIREMFNGEFRFLCRMT------EETERLCIEAVFGIV 283

Qy     278 GFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEKCNQPTP 337
           | | |  |      |   |  || : : : |: | : ||    : : |  ||::      |
Db     284 GRNWNEFNRKMFPVILGHYPAGVAAKQVKHFIQIIKSGRFAPYSY-SSNKNMQLYRDHLP 342

Qy     338 VRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIP--EWAHVDFIWGLDA 395
           || :  :|||| ::    | | :|:||: :  :: |:     :|  |: |:||:| :|
Db     343 PRYNLSLVTVPTFVYYSTNDLLCHPKDVESMCDDLGNVTGKYLVPQKEFNHMDFLWAIDV 402

Qy     396 PHRMYNEIIHLM------QQEETNLSQ 416
             :|  :: ::        || | |:
Db     403 RKMLYRRMLQVLGKVPEGSPEEANRSR 429


RESULT   8
TGL1_YEAST
ID   TGL1_YEAST     STANDARD;      PRT;   548 AA.
AC   P34163;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   TRIGLYCERIDE LIPASE-CHOLESTEROL ESTERASE (EC 3.1.1.-).
GN   TGL1 OR YKL140W OR YKL5.
OS   Saccharomyces cerevisiae (Baker's yeast).
```

```
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92245761; PubMed=1574929;
RA   Abraham P.R., Mulder A., Van'T Riet J., Planta R.J., Raue H.A.;
RT   "Molecular cloning and physical analysis of an 8.2 kb segment of
RT   chromosome XI of Saccharomyces cerevisiae reveals five tightly linked
RT   genes.";
RL   Yeast 8:227-238(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rad M.R., Xu G., Kirchrath L., Fritz C., Keuchel H., Hollenberg C.P.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z25464; CAA80958.1; -.
DR   EMBL; Z28140; CAA81981.1; -.
DR   PIR; S37969; S37969.
DR   PIR; S39000; S39000.
DR   SGD; S0001623; TGL1.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   Pfam; PF00561; abhydrolase; 1.
KW   Hydrolase; Lipid degradation.
SQ   SEQUENCE   548 AA;  62979 MW;  32D1F230701CB083 CRC64;


  Query Match             18.8%;  Score 422.5;  DB 1;  Length 548;
  Best Local Similarity   31.6%;  Pred. No. 6.7e-29;
  Matches  119;  Conservative   66;  Mismatches  127;  Indels   65;  Gaps   15;

Qy      61 EEYEVATEDGYILSVNRIPRGLVQP--KKTGSRPVVLLQHGLVGGASNWISNLP-NNSLG 117
           |:: | ||| |||:::|||     |  |   :  || | |||:  :  |  |:  : :|
Db      78 EDHLVRTEDNYILTLHRIP-----PISKNRFNNKVVYLHHGLLMCSDVWCCNIERHKNLP 132

Qy     118 FILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQ 177
           |:| | |:||||||:||| :|  |        ::|| || || | ||:|  | |||  |
Db     133 FVLHDLGYDVWMGNNRGNKYSTAHLNKPPKSNKFWDFSIDEFAFFDIPNSIEFILDITKV 192

Qy     178 EKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVK--HAK--------SPGTK 227
           :|:  :|:|||:   | |||   :| :|:  : |:||  | |  | :         |||
Db     193 DKVICIGFSQGSAQMFAAFSLSEKLNRKVSHFIAIAPAMTPKGLHNRIVDTLAKSSPGFM 252

Qy     228 FLLLPDMMIKGLFGKKEFLYQT----RFLRQLVIYLCGQVILDQICSNIMLLLGGFNTNN 283
           :|          ||:|  |       | |   :  ||    :| | : |:     ||  :
Db     253 YL---------FFGRKIVLPSAVIWQRTLHPTLFNLC----ID-IANKIL-----FNWKS 293

Qy     284 MNM-------SRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFD-----WGSETKNLEK 331
            |:       | | :|:       |||::|:|| | : | : : |:       | |: :
Db     294 FNILPRQKIASYAKLYST-----TSVKSIVHWFQILRSQKFQMFEESDNMLNSLTRPYQI 348

Qy     332 CNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIW 391
             | ||       :: :|  :  || | | : : :|  |    | ::   :  : |:| ||
Db     349 ANFPTRT-----NIKIPILLIYGGIDSLVDIDVMKKNLP--FNSVFDVKVDNYEHLDLIW 401

Qy     392 GLDAPHRMYNEIIHLMQ 408
           | ||  :  :::  ::
Db     402 GKDADTLVIAKVLRFIE 418


RESULT   9
YD94_METJA
ID   YD94_METJA     STANDARD;      PRT;   987 AA.
AC   Q58789;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ1394.
GN   MJ1394.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: TO M.JANNASCHII MJ1393 AND A.FULGIDUS AF2028.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67579; AAB99404.1; -.
DR   TIGR; MJ1394; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     12     32       POTENTIAL.
FT   TRANSMEM    958    978       POTENTIAL.
SQ   SEQUENCE   987 AA;  112360 MW;  D1E628FFB28CA86D CRC64;


  Query Match              5.6%;  Score 126;  DB 1;  Length 987;
  Best Local Similarity   23.1%;  Pred. No. 0.0072;
  Matches  105;  Conservative   56;  Mismatches  149;  Indels  144;  Gaps   29;

Qy       3 ETLSRQWIVSHRMEMWLLILVAYMFQR--------NVNSVHMPTKAVDPEAFMN------ 48
           ||     | :::|  | : |: |: :          ||| :   :     : | |
Db     340 ETYDVTKIPANKMSEWTVNLIVYLNKNLVPNGENVNVNVIKYLSNGQFYDNFENWTGWNQ 399

Qy      49 -------ISEIIQHQG-YPCEEYEVATEDGYILSVNRIPRG--LVQPKKTGSRPVVLLQH 98
                | |  | | |  |:| ::|      |:|  | |   : ||: | | ||:
Db     400 YKNGIVQWSSIQSHSGNYSLEKYGIST------SLNNDPNGGYKLLPKEIG-RDVVI--S 450

Qy      99 GLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKT--LSIDQ-------- 148
           | |   |||    | : :|  | |  ||       | ::   | :  :|||:
Db     451 GWVYRPSNW-GGGPIDRIG--LEDENFD------GYSFEVNHYSNYISIDRRTNGNPTE 500

Qy     149 ---DEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQK 205
               : :|    ||   |:|              ||    || ||    | |||
Db     501 ISPEVYWNPPEDEWYYFEL-------------KI----YSNGT----ITFST------- 531

Qy     206 IKMYFALAPI-ATVKHAKSPGTKF----------LLLPDMMIKG----LFGKKEFLYQTR 250
             |:    : |||    :  |||            : |: :       :| | : |
Db     532 ---YYQNGSLAATVSTIDNTYTKFDRVVIHGGYVYYVDDLEVNSKNFDFYGDKNWKYL-- 586

Qy     251 FLRQLVIYLCGQVILDQICSNIMLLLGG------FNTNNMNMSRASVYAAHTL-----AG 299
                  ::          :|  |      :||:|:|   |  :   ||     :
Db     587 -----------EITSANSSEGTAVLFDGDYFKKDYNTSNLN---AINWTNITLNWSNDSA 632

Qy     300 TSVQNIL-HWSQAVNSGELRAFDWGSETKNLEKCNQPTPVR--YRVRDMTVPTAMWTGGQ 356
           | | |:| ::| :     |  : :     |  |  | ::  |     :: |   | |:
```

```
Db    633 TLVFNVLGNYSYSERDNILAKYGFAKILFNYNGTNTNTSIKGVYASGSYSISTDHGTTGE 692

Qy    357 --DWLSN---PEDVKMLLSEVTNLIYHKNIPEWA 385
Db    693 INIWIENVTFKNDAKSYSFNLTNL----NI--WA 720


RESULT 10
ORC2_CAEEL
ID   ORC2_CAEEL     STANDARD;      PRT;   430 AA.
AC   Q21037;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PROBABLE ORIGIN RECOGNITION COMPLEX SUBUNIT 2.
GN   F59E10.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Swinburne J.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
RA   Jones S.J.M.;
RL   Submitted (MAR-1997) to the SWISS-PROT data bank.
CC   -!- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT
CC       BINDS ORIGINS OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL
CC       REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO
CC       THE ARS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION IN AN
CC       ATP-DEPENDENT MANNER.
CC   -!- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE ORC2 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z36949; CAA85415.1; -.
DR   WormPep; F59E10.1; CE11492.
KW   DNA replication; Nuclear protein.
SQ   SEQUENCE   430 AA;  49319 MW;  54FC086BB4AD9670 CRC64;


  Query Match             4.7%;  Score 105;  DB 1;  Length 430;
  Best Local Similarity   21.9%;  Pred. No. 0.15;
  Matches   61;  Conservative   42;  Mismatches   87;  Indels   88;  Gaps   13;

Qy     85 PKKTGSRPVVLLQHGLVGGASNWISNLPNN------SLGF-----ILADAGFDVW--MGN 131
Db     25 PEKEGSR------QKKTNGKENASRNLQSNLEEDLEQLGFEDETVSMAQSAIENYFMQGK 78

Qy    132 S----RGNAWSRKHK------TLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIY 181
Db     79 SASERMNNAKSRRGRRAGNGNTEEIEEDDEISNAITDFTKCDLPGLRNYITKKDNTEFEK 138

Qy    182 YVGYSQGTTMGFIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFG 241
Db    139 RLEHLADNDFG-----------KWKLYLAAG----------------FNILLHGVGS 168

Qy    242 KKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGGFNTN---NMNMSRASVYAAHTLA 298
Db    169 KRDVL--TEFENELSDYTYMRVDARKDGLNVKVLLGAINENMKLNCNVKRG--------- 217

Qy    299 GTSVQNILHWSQAV----NSGEL-------RAFDWGSE 325

              |: : |::::    || :|       | || |:
Db    218 ----QSTISWARSIRRKMNSQQLILIIDNIEAPDWRSD 251


RESULT 11
PAFA_CAVPO
ID   PAFA_CAVPO     STANDARD;      PRT;   436 AA.
AC   P70683;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE PRECURSOR (EC 3.1.1.47)
DE   (PAF ACETYLHYDROLASE) (PAF 2-ACYLHYDROLASE) (LDL-ASSOCIATED
DE   PHOSPHOLIPASE A2) (LDL-PLA(2)) (2-ACETYL-1-ALKYLGLYCEROPHOSPHOCHOLINE
DE   ESTERASE) (1-ALKYL-2-ACETYLGLYCEROPHOSPHOCHOLINE ESTERASE).
GN   PLA2G7 OR PAFAH.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HARTLEY; TISSUE=Liver;
RX   MEDLINE=97103479; PubMed=8947850;
RA   Karasawa K., Kuge O., Kawasaki K., Nishijima M., Nakano Y., Tomita M.,
RA   Yokoyama K., Setaka M., Nojima S.;
RT   "Cloning, expression and characterization of plasma
RT   platelet-activating factor-acetylhydrolase from guinea pig.";
RL   J. Biochem. 120:838-844(1996).
CC   -!- FUNCTION: MODULATES THE ACTION OF PLATELET-ACTIVATING FACTOR (PAF)
CC       BY HYDROLYZING THE SN-2 ESTER BOND TO YIELD THE BIOLOGICALLY
CC       INACTIVE LYSO-PAF. HAS A SPECIFICITY FOR SUBSTRATES WITH A SHORT
CC       RESIDUE AT THE SN-2 POSITION. IT IS INACTIVE AGAINST LONG-CHAIN
CC       PHOSPHOLIPIDS.
CC   -!- CATALYTIC ACTIVITY: 2-ACETYL-1-ALKYL-SN-GLYCERO-3-PHOSPHOCHOLINE +
CC       H(2)O = 1-ALKYL-SN-GLYCERO-3-PHOSPHOCHOLINE + ACETATE.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC       HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D67037; BAA11054.1; -.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   PROSITE; PS00120; LIPASE_SER; 1.
KW   Hydrolase; Lipid degradation; Glycoprotein; Signal.
FT   SIGNAL       1     21       BY SIMILARITY.
FT   CHAIN       22    436       PLATELET-ACTIVATING FACTOR
FT                               ACETYLHYDROLASE.
FT   ACT_SITE    271    271       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    294    294       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    349    349       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     76     76       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    200    200       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    324    324       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   436 AA;  49062 MW;  C359D96E392FFE11 CRC64;


  Query Match             4.5%;  Score 101;  DB 1;  Length 436;
  Best Local Similarity   20.4%;  Pred. No. 0.35;
  Matches   85;  Conservative   60;  Mismatches  149;  Indels  122;  Gaps   21;

Qy     21 ILVAYMFQRNVNSVHMPTKAVD-PEAFMNISEIIQHQGYPCEEY-EVATED------GY 71
Db     70 LMSGYTNQSSFLRLYYPSQDNDFPDALW----------IPNEEYFQGLTETLGASSFLGK 119
```

```
Qy      72 ILSV----NRIPRGLVQPKKTGSR-PVVLLQHGLVGGASNWISNLPNNSLGFILADAGFD 126
           :| :     ::|     | ||| : |::: |||  ||   |     :::|  ||  ||
Db     120 LLKLLYGSVKVPAKWNSPLKTGEKYPLIIFSHGL--GAFRSI----YSAIGIELASHGFI 173

Qy     127 V---------------------------WMGNSRGNAWSRKHKTLSIDQDE------FW 152
           |                           |:    ||  : : |     : |       |
Db     174 VAAVEHRDESAAATYYFQDAPAAESGNRSWIYYKVGNLETEERKRQLRQRGEECSQALSW 233

Qy     153 AFSYDEMARFDLPAVINFILQK----TGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKM 208
           | ||          :|| :|:      : |:  :|:| |   |     |: |  |: :
Db     234 LLSIDEGEPVKNVLDLNFDIQQLKGSLDRSKVAIIGHSFG---GATVIQTLSE-DQRFRC 289

Qy     209 YFAL----APIATVKHAKSPGTKFLLLPDMMIKGLFGKK-EFLYQTRFLRQLV------- 256
             ||     |:    |:| |   | :  :         || |  || :  |:::
Db     290 GIALDPWMFPVGEDVHSKIPQPLFFINSEYFQSANDTKKIEKFYQPQKERKMIAVKGSVH 349

Qy     257 ------IYLCGQVILDQICSNIMLLLGGFNTNNMNM---SRASVYAAHTLAGTSVQNILH 307
                 :  |::|        || | |   : : |   ::||:       |   :|
Db     350 HNFVDFTFATGKII------GQMLSLKGKIDSEVAMDLINKASLAFLQKYLGLD-KNFDQ 402

Qy     308 WSQAVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPE 363
           |:  :          : :||          : : |:|| | :       ||:
Db     403 WNSLME----------GDDENL-----------IPEFTIPTTMQSSTGTEQRNPD 436


RESULT  12
Y193_HAEIN
ID   Y193_HAEIN     STANDARD;      PRT;   287 AA.
AC   Q57427; O05013;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PUTATIVE ESTERASE/LIPASE HI0193 (EC 3.1.-.-).
GN   HI0193.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- SIMILARITY: BELONGS TO A FAMILY OF ESTERASES THAT GROUPS TOGETHER
CC       PSEUDOMONA TROPINESTERASE, DMPD, TODF AND XYLF.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32704; AAC21862.1; -.
DR   TIGR; HI0193; -.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   Pfam; PF00561; abhydrolase; 1.
KW   Hypothetical protein; Hydrolase; Serine esterase; Complete proteome.
FT   ACT_SITE    119    119       BY SIMILARITY.
FT   ACT_SITE    266    266       BY SIMILARITY.
SQ   SEQUENCE   287 AA;  32987 MW;  F2B548619C060619 CRC64;


  Query Match             4.3%;  Score 96.5;  DB 1;  Length 287;
  Best Local Similarity   19.1%;  Pred. No. 0.48;
  Matches   62;  Conservative   39;  Mismatches  119;  Indels  105;  Gaps   12;

Qy      84 QPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKT 143
           | |:| : ||::  ||| |   |          || |                |:|  :
Db      39 QVKQTINTPVLIFIHGLFGDMDN---------LGVI--------------ARAFSEHYSI 75

Qy     144 LSIDQDEFWAFSYDEMARFDLPAV-INFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPEL 202
           | ||        : |   : | |  :  :::      |:  :|:| |         :  |||
Db      76 LRIDLRNHGHSFHSEKMNYQLMAEDVIAVIRHLNLSKVILIGHSMGGKTAMKITALCPEL 135

Qy     203 AQKIKMYFALAPIATVKHAKSPGTKFLLLPDM------------MIKGLFGKKEFLYQTR 250
           :|                          |:: ||          :  |||  |    :
Db     136 VEK----------------------LIVIDMSPMPYEGFGHKDVFNGLFAVKNAKPEN- 171

Qy     251 FLRQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQ 310
             ||    :  | | |:    :  :|  |: |: :  | ::    |    :  ||: |
Db     172 --RQQAKPILKQEINDE--DVVQFMLKSFDVNSADCFRFNL----TALFNNYANIMDWE- 222

Qy     311 AVNSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLS 370
                 ::| |                                  ||    ||       |: : :|
Db     223 -----KVRVF-------------------------TPTLFIKGGNSSYIKIENSEKILE 251

Qy     371 EVTNLIYH--KNIPEWAHV---DFI 390
           :  |           | |    ||:
Db     252 QFPNATAFTINGSGHWVHAEKPDFV 276


RESULT  13
LIN1_HUMAN
ID   LIN1_HUMAN     STANDARD;      PRT;  1259 AA.
AC   P08547;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   01-AUG-1988 (Rel. 08, Last annotation update)
DE   LINE-1 REVERSE TRANSCRIPTASE HOMOLOG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=86230917; PubMed=2423883;
RA   Hattori M., Kuhara S., Takenaka O., Sakaki Y.;
RT   "L1 family of repetitive DNA sequences in primates may be derived
RT   from a sequence encoding a reverse transcriptase-related protein.";
RL   Nature 321:625-628(1986).
CC   -!- MISCELLANEOUS: THIS SEQUENCE WAS CONSTRUCTED FROM AN ALIGNMENT OF
CC       PUBLISHED AND UNPUBLISHED SEQUENCES, DETERMINED IN VARIOUS
CC       LABORATORIES, BELONGING TO THE LINE-1 FAMILY.
DR   PIR; A25313; GNHUL1.
DR   InterPro; IPR000097; AP_endonclse_family_1.
DR   InterPro; IPR000300; IPPc.
DR   InterPro; IPR000477; RVTse.
DR   Pfam; PF01260; AP_endonucleas1; 1.
DR   Pfam; PF00078; rvt; 1.
DR   SMART; SM00128; IPPc; 1.
KW   RNA-directed DNA polymerase.
SQ   SEQUENCE   1259 AA;  147112 MW;  F3BC18A3803919F5 CRC64;


  Query Match             4.3%;  Score 95.5;  DB 1;  Length 1259;
  Best Local Similarity   19.7%;  Pred. No. 4.4;
  Matches   90;  Conservative   67;  Mismatches  170;  Indels  129;  Gaps   23;

Qy      28 QRNVNSVHMPTKAVDPEAFMNI---SEIIQHQGYPCEEYE------------------- 64
           |  | |:: |  : : || :|     :    :|:  | |:
Db     442 QEEVESLNRPITSSEIEAIINSLPNKKSPGPEGFTAEFYQRYKEELVPFLLKLFQSIEKE 501
```

```
Qy      65 -VATEDGYILSVNRIPR-GLVQPKKTGSRPVVL-------LQHGLVGGASNWISNL-PNN 114
Db     502 GILPNSFYEASIILIPKPGRDTTKKENFRPISLMNIDAKILNKILANQIQQHIKKLIHHD 561

Qy     115 SLGFILADAGFDVWMGNSR--------GNAWSRKHKTLSIDQDEFWAFSYDEM------- 159
Db     562 QVGFIPAMQG---WFNIRKSINIIQHINRTKDTNHMIISIDAEK----AFDKIQQPFMLK 614

Qy     160 ----------------ARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELA 203
Db     615 PLNKLGIDGTYLKIIRAIYDKPTA-NIIL--NGQ-KLEAPPLKTGTRQGCPLSPLLPNIV 670

Qy     204 QKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQV 263
Db     671 LEV-----LA--RAIRQEKE-----------IKGIQLGKEEVKLSLFADDMIVYLENPI 711

Qy     264 ILDQICSNIMLLLGGFN-----TNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELR 318
Db     712 VSAQ---NLLKLISNFSKVSGYKINVQKSQAFLYTNNRQTESQIMSELPFTIASKRIKYL 768

Qy     319 AFDWGSETKNLEKCNQPTPVRYRVRDMT-----VPTAMWTGGQDWLSNPEDVKMLLSEVT 373
Db     769 GIQLTRDVKDLFKENY-KPLLNEIKEDTNKWKNIPCS-------WVGRINIVKMAI--LP 818

Qy     374 NLIYHKN-IP----------EWAHVDFIWGLDAPH 397
Db     819 KVIYRFNAIPIKLPMTFFTELEKTTLKFIWNQKRAH 854
```

```
RESULT  14
PRXC_STRAU
ID   PRXC_STRAU     STANDARD;      PRT;   278 AA.
AC   O31168;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   NON-HEME CHLOROPEROXIDASE (EC 1.11.1.10) (CHLORIDE PEROXIDASE)
DE   (CPO-T) (CHLOROPEROXIDASE T).
GN   CPO OR CPOT.
OS   Streptomyces aureofaciens.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1894;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TUE24;
RA   Pelletier I., Altenbuchner J., van Pee K.-H.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).
RC   STRAIN=TUE24;
RX   MEDLINE=98307994; PubMed=9642069;
RA   Hofmann B., Tolzer S., Pelletier I., Altenbuchner J., van Pee K.-H.,
RA   Hecht H.-J.;
RT   "Structural investigation of the cofactor-free chloroperoxidases.";
RL   J. Mol. Biol. 279:889-900(1998).
CC   -!- CATALYTIC ACTIVITY: 2 RH + 2 CHLORIDE + H(2)O(2) = 2 RCL + 2
CC       H(2)O.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SIMILARITY: TO OTHER BACTERIAL NON-HEME BROMO- AND CHLORO-
CC       PEROXIDASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF031242; AAB86626.1; -.
DR   PDB; 1A7U; 17-JUN-98.
DR   PDB; 1A8U; 17-JUN-98.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   Pfam; PF00561; abhydrolase; 1.
KW   Oxidoreductase; Peroxidase; 3D-structure.
FT   ACT_SITE     99     99       BY SIMILARITY.
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   ACT_SITE    258    258       BY SIMILARITY.
SQ   SEQUENCE   278 AA;  30354 MW;  8C97A87251FBEDE5 CRC64;
```

```
  Query Match             4.2%;  Score 93.5;  DB 1;  Length 278;
  Best Local Similarity   22.6%;  Pred. No. 0.84;
  Matches   72;  Conservative   36;  Mismatches  114;  Indels   97;  Gaps   17;

Qy      91 RPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDE 150
Db      25 QPVVLI-HGFPLSGHSW------ERQSAALLDAGYRVITYDRRGFGQSSQPTT------- 70

Qy     151 FWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYF 210
Db      71 --GYDYDTFA-----ADLNTVLETLDLQDAVLVGFSMGT----------GEVARYVSSY- 112

Qy     211 ALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICS 270
Db     113 GTARIAKVAFLASL-EPFLLKTDDNPDGA-APKEF------------------FDGIVA 151

Qy     271 NIMLLLGGFNTNNMNMSRASVYAAHTLAGT-----SVQNILHWSQAVNSGELRA----FD 321
Db     152 AVKADRYAFYTGFFN----DFYNLDENLGTRISEEAVRN--SWNTAASGGFFAAAAAPTT 205

Qy     322 WGSETK-NLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKN 380
Db     206 WYTDFRADIPRIDVPALILHGTGDRTLPI--------------------ENTARVFHKA 244

Qy     381 IP--EWAHVDFIWGLDAPH 397
Db     245 LPSAEYVEVE-----GAPH 258
```

```
RESULT  15
PYRD_PLAFA
ID   PYRD_PLAFA     STANDARD;      PRT;   569 AA.
AC   Q08210;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   DIHYDROOROTATE DEHYDROGENASE HOMOLOG PRECURSOR (EC 1.3.3.1)
DE   (DIHYDROOROTATE OXIDASE).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=94049995; PubMed=8232427;
RA   Leblanc S.B., Wilson C.M.;
RT   "The dihydroorotate dehydrogenase gene homologue of Plasmodium
RT   falciparum.";
RL   Mol. Biochem. Parasitol. 60:349-352(1993).
CC   -!- CATALYTIC ACTIVITY: L-DIHYDROOROTATE + O(2) = OROTATE + H(2)O(2).
CC   -!- COFACTOR: FAD.
CC   -!- PATHWAY: FOURTH STEP IN PYRIMIDINE BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; L15446; AAC37170.1; -.
DR   InterPro; IPR001295; DHO_dh.
DR   InterPro; IPR003009; FMN_enzyme.
DR   Pfam; PF01180; DHOdehase; 1.
DR   PROSITE; PS00911; DHODEHASE_1; 1.
DR   PROSITE; PS00912; DHODEHASE_2; 1.
KW   Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FAD;
KW   Transit peptide; Mitochondrion; Inner membrane.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?    569       DIHYDROOROTATE DEHYDROGENASE HOMOLOG.
FT   NP_BIND     235    243       FAD (NAD PART) (POTENTIAL).
SQ   SEQUENCE   569 AA;  65558 MW;  88880384EBD52FE3 CRC64;


  Query Match             4.1%;  Score 93;  DB 1;  Length 569;
  Best Local Similarity   28.8%;  Pred. No. 2.5;
  Matches   30;  Conservative   14;  Mismatches   52;  Indels    8;  Gaps    1;

Qy     218 VKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLG 277
           :|  ||   | : |  : : ||:|  |      ||  : :  | : |  :|| :: ||||
Db     133 IKKEKSKHKKIIFLLFVSLFGLYGFFESYNPEFFLYDIFLKFCLKYIDGEICHDLFLLLG 192

Qy     278 GFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFD 321
             :|    : |  |:||         || |       |    ||
Db     193 KYNILPYDTSNDSIYAC--------TNIKHLDFINPFGVAAGFD 228


Search completed: January 30, 2002, 23:10:17
Job time: 533 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 30, 2002, 23:00:04 ; Search time 91.96 Seconds
(without alignments)
672.827 Million cell updates/sec

Title: US-09-578-063-47
Perfect score: 2247
Sequence: 1 MLETLSRQWIVSHRMEMWLL.....IHLMQQEETNLSQGRCEAVL 423

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1290	57.4	399	4	Q16529	Q16529 homo sapien
2	1220.5	54.3	397	11	Q64194	Q64194 rattus sp.
3	1197.5	53.3	397	11	Q9Z0M5	Q9z0m5 mus musculu
4	1134.5	50.5	395	11	Q9D798	Q9d798 mus musculu
5	1134.5	50.5	395	11	Q9D6X0	Q9d6x0 mus musculu
6	1134.5	50.5	395	11	Q9D6Q6	Q9d6q6 mus musculu
7	1134.5	50.5	395	11	Q9D6P3	Q9d6p3 mus musculu
8	1133.5	50.4	395	11	Q9CPP8	Q9cpp8 mus musculu
9	1133.5	50.4	395	11	Q9CPP7	Q9cpp7 mus musculu
10	1131.5	50.4	395	11	Q9D6L9	Q9d619 mus musculu
11	1129.5	50.3	395	11	Q9D6T5	Q9d6t5 mus musculu
12	1128.5	50.2	395	11	Q9D767	Q9d767 mus musculu
13	1127.5	50.2	395	11	Q9D7C5	Q9d7c5 mus musculu
14	1126.5	50.1	395	11	Q9D766	Q9d766 mus musculu
15	1126.5	50.1	395	11	Q9D6L1	Q9d611 mus musculu
16	1125.5	50.1	395	11	Q9D6Q3	Q9d6q3 mus musculu
17	1124.5	50.0	395	11	Q9D760	Q9d760 mus musculu
18	1124.5	50.0	395	11	Q9D6S5	Q9d6s5 mus musculu
19	1122.5	50.0	395	11	Q9D796	Q9d796 mus musculu
20	1119.5	49.8	395	11	Q9D6N8	Q9d6n8 mus musculu
21	1116.5	49.7	395	11	Q9D6Q2	Q9d6q2 mus musculu
22	1115.5	49.6	395	11	Q9D6M9	Q9d6m9 mus musculu
23	752.5	33.5	403	5	O61866	O61866 caenorhabdi
24	751	33.4	411	5	Q20449	Q20449 caenorhabdi
25	730	32.5	405	5	Q93789	Q93789 caenorhabdi
26	715	31.8	169	11	Q9D2L7	Q9d217 mus musculu
27	707.5	31.5	405	5	O16956	O16956 caenorhabdi
28	692	30.8	411	5	Q94252	Q94252 caenorhabdi
29	655.5	29.2	426	5	O17766	O17766 caenorhabdi
30	655.5	29.2	434	5	Q9VKT9	Q9vkt9 drosophila
31	639	28.4	456	5	Q9VKS5	Q9vks5 drosophila
32	636	28.3	398	5	Q9V796	Q9v796 drosophila
33	625.5	27.8	616	5	O77107	O77107 plodia inte
34	588.5	26.2	838	5	Q9VBK6	Q9vbk6 drosophila
35	574.5	25.6	559	5	Q17219	Q17219 bombyx mori
36	567.5	25.3	457	5	Q9VKT2	Q9vkt2 drosophila
37	559	24.9	399	5	Q9VPE9	Q9vpe9 drosophila
38	535	23.8	416	5	Q9VQQ5	Q9vqq5 drosophila
39	529.5	23.6	406	5	Q9VKT7	Q9vkt7 drosophila
40	527	23.5	504	5	Q94568	Q94568 galleria me
41	505	22.5	443	3	P78898	P78898 schizosacch
42	502	22.3	435	5	Q9VG46	Q9vg46 drosophila
43	488	21.7	355	5	Q9VKT0	Q9vkt0 drosophila
44	485	21.6	356	5	Q9VKS9	Q9vks9 drosophila
45	478.5	21.3	401	5	Q9U276	Q9u276 caenorhabdi

ALIGNMENTS

```
RESULT   1
Q16529
ID   Q16529      PRELIMINARY;      PRT;   399 AA.
AC   Q16529;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LYSOSOMAL ACID LIPASE PRECURSOR.
GN   LAL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Du H., Gregory G.A.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; Z31690; CAA83495.1; -.
DR   EMBL; U08464; AAB60328.1; -.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
KW   Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    399       LYSOSOMAL ACID LIPASE.
FT   VARIANT      16     16       T -> P (IN REF. 1).
FT   VARIANT      23     23       R -> G (IN REF. 1).
SQ   SEQUENCE   399 AA;  45518 MW;  AE051EE098A8152B CRC64;


  Query Match             57.4%;  Score 1290;  DB 4;  Length 399;
  Best Local Similarity   61.0%;  Pred. No. 2.8e-106;
  Matches  242;  Conservative   59;  Mismatches   94;  Indels    2;

Qy      15 MEMWLLILVAYMFQRNVNSVHMPTK--AVDPEAFMNISEIIQHQGYPCEEYEVATEDG
           |:|  | ||  :    ::|     |  ||||| ||:|||| : |:| ||| | ||||
Db       1 MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYWGFPSEEYLVETEDG
```

```
Qy      73 LSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNS 132
           | :|||| |       | :||| |||||:  :|||::|| |:||||||||||||||||||
Db      61 LCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANSSLGFILADAGFDVWMGNS 120

Qy     133 RGNAWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMG 192
           ||| |||||||||: ||||||||||||||::|||| ||||| |||||::||||:|||||:|
Db     121 RGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFILNKTGQEQVYYVGHSQGTTIG 180

Qy     193 FIAFSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFL 252
           ||||| :||||::|||:||| |:|:|    ||  |   ||| :|| ||| |||| |: ||
Db     181 FIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDLFGDKEFLPQSAFL 240

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
           : |  ::|  ||| ::| |:  || ||| |:|||| ||  |: |||||||:|||||||
Db     241 KWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSRVDVYTTHSPAGTSVQNMLHWSQAV 300

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
              : ||||||  ||    ||  |  | |:|| ||||:|:|| |||::  || :||:::
Db     301 KFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDMLVPTAVWSGGHDWLADVYDVNILLTQI 360

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQ 409
           |||::|::|||| |:|||||||| |:||:||::||::
Db     361 TNLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRK 397


RESULT   2
Q64194
ID   Q64194      PRELIMINARY;      PRT;   397 AA.
AC   Q64194;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LYSOSOMAL ACID LIPASE.
OS   Rattus sp.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10118;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96129534; PubMed=8576647;
RA   Nakagawa H., Matsubara S., Kuriyama M., Yoshidome H., Fujiyama J.,
RA   Yoshida H., Osame M.;
RT   "Cloning of rat lysosomal acid lipase cDNA and identification of the
RT   mutation in the rat model of Wolman's disease.";
RL   J. Lipid Res. 36:2212-2218(1995).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; S81497; AAB36043.2; -.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   Pfam; PF00561; abhydrolase; 1.
SQ   SEQUENCE   397 AA;  45186 MW;  97A38595A0523947 CRC64;


  Query Match             54.3%;  Score 1220.5;  DB 11;  Length 397;
  Best Local Similarity   58.2%;  Pred. No. 4.1e-100;
  Matches  220;  Conservative   67;  Mismatches   84;  Indels    7;  Gaps    3;

Qy      37 PT---KAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIPRGLVQPKKTGSRPV 93
           ||    |||||| ||::||| | |||  |: | | ||||| |:||| |       | :||
Db      20 PTGTISAVDPEANMNVTEIIMHWGYP--EHSVQTGDGYILGVHRIPHGRKNQFDKGPKPV 77

Qy      94 VLLQ--HGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRKHKTLSIDQDEF 151
           | ||  || :  :|||::|: ||||||||||||||||||||| |||||||||: |||:
Db      78 VYLQWRHGFLADSSNWVTNIDNNSLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEY 137

Qy     152 WAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYFA 211
           ||||:|||||::|||| ||:|| |||||::| ||:||| |:|||||| |||||:|:||:||
Db     138 WAFSFDEMAKYDLPASINYILNKTGQEQLYNVGHSQGCTIGFIAFSQMPELAKKVKMFFA 197

Qy     212 LAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSN 271
           |||: ::  |  |  |   |||:::: |||:|:|| |:  :: |  ::|  ||: ::|:|
Db     198 LAPVLSLNFASGPMVKLGRLPDLLLEDLFGQKQFLPQSAMVKWLSTHICTHVIMKELCAN 257

Qy     272 IMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGSETKNLEK 331
           |  |: |||  |:|||  ||  |  |||||||::||:| |   :|:|||||  ||
Db     258 IFFLICGFNEKNLNMSRVDVYTTHCPAGTSVQNMVHWTQVVKYHKLQAFDWGSSDKNYFH 317

Qy     332 CNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHKNIPEWAHVDFIW 391
           ||  |  | ::|| :|||:|:||:|||::  |: :||:|:  |:|||||||| |:||||
Db     318 YNQSYPPLYSIKDMQLPTALWSGGKDWLADTSDINILLTEIPTLVYHKNIPEWDHLDFIW 377

Qy     392 GLDAPHRMYNEIIHLMQQ 409
           ||||| |:|||:: ||::
Db     378 GLDAPWRLYNEVVSLMKK 395


RESULT   3
Q9Z0M5
ID   Q9Z0M5      PRELIMINARY;      PRT;   397 AA.
AC   Q9Z0M5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LYSOSOMAL ACID LIPASE PRECURSOR.
GN   LIP1 OR LAL.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57 BLACK/6X CBA; TISSUE=LIVER;
RX   MEDLINE=96363957; PubMed=8725147;
RA   Du H., Witte D.P., Grabowski G.A.;
RT   "Tissue and cellular specific expression of murine lysosomal acid
RT   lipase mRNA and protein.";
RL   J. Lipid Res. 37:937-949(1996).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; Z31689; CAA83494.1; -.
DR   MGD; MGI:96789; Lip1.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    397       LYSOSOMAL ACID LIPASE.
SQ   SEQUENCE   397 AA;  45550 MW;  F886C39E1CCFA91F CRC64;


  Query Match             53.3%;  Score 1197.5;  DB 11;  Length 397;
  Best Local Similarity   55.6%;  Pred. No. 4.5e-98;
  Matches  217;  Conservative   70;  Mismatches   100;  Indels    3;  Gaps    1;

Qy      23 VAYMFQRNVNSVHMPT---KAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSVNRIP 79
           : ::|   :    :||    |||||  ||::|||   ||| ||: | | |||||||:|||
Db       6 LVFVFTIGILLSRVPTGTVSAVDPEVNMNVTEIIMRWGYPGEEHSVLTGDGYILSIHRIP 65

Qy      80 RGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSR 139
           ||       | |||| |||||:  :|||::|: |:||||:||| ||||||||||| ||
Db      66 RGWKNHFGKGPRPVVYLQHGLLADSSNWVTNIDNSSLGFLLADRGFDVWMGNSRGNTWSL 125

Qy     140 KHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTM 199
           ||||||: |||||||:||||::|||| ||:|| |||||:|||||:||| |:|||||| |
Db     126 KHKTLSVSQDEFWAFSFDEMAKYDLPASINYILNKTGQEQIYYVGHSQGCTIGFIAFSQM 185

Qy     200 PELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYL 259
           ||||:||||:  |||: ::  |  |  :   ||| ::| :||:|:|| |:  |: | |::
Db     186 PELAKKIKMFLVLAPVLSLNFASGPLLQLGRLPDPLLKDMFGQKQFLPQSAMLKWLSIHV 245

Qy     260 CGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAVNSGELRA 319
           | ||: ::|:|:  || |||  |:|||| ||  |  |   |||:||| |     :|:|
Db     246 CTHVIMKELCANVFFLLCGFNEKNLNMSRVDVYTTHCPAELLVQNMLHWGQVFKYRKLQA 305

Qy     320 FDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHK 379
```

```
          |||||  ||    ||  |  | :::| :|||:|:||:|||::  |: :||::: |:|||
Db    306 FDWGSSEKNYFHYNQSFPPSYNIKNMRLPTALWSGGRDWLADINDITILLTQIPKLVYHK 365

Qy    380 NIPEWAHVDFIWGLDAPHRMYNEIIHLMQQ 409
          ||||| |:|||||||| ::|:||| ||::
Db    366 NIPEWDHLDFIWGLDAPWKLYDEIISLMKK 395


RESULT   4
Q9D798
ID   Q9D798      PRELIMINARY;      PRT;   395 AA.
AC   Q9D798;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2310051B21RIK PROTEIN.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK009431; BAB26283.1; -.
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44603 MW;  D3FD8B6FEA671E3E CRC64;


  Query Match             50.5%;  Score 1134.5;  DB 11;  Length 395;
  Best Local Similarity   53.5%;  Pred. No. 1.8e-92;
  Matches  213;  Conservative   70;  Mismatches  108;  Indels    7;  Gaps    3;

Qy     17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
          ||||:: : :     : :        :||| ||:|::| : ||| ||||| ||||||| |
Db      1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy     76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
          ||| |    :  | |||  |||||:  |:|||:||||||| |||||||:|||:||||||
Db     61 YRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy    136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
          ||||:   | |  ||||||:||||::|||| |:||:||||||||:|||:|||||:||||
Db    121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy    196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
          |||  | ||:||| ::||||:||||: :||  |  |:|  ::| :|| | |:   |  :||
Db    181 FSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPKFLLKVIFGNKMFMPHNYLDQFL 240

Qy    253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
          .  |  | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db    241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRLDVYLGHNPAGTSTQDLFHWAQLA 297

Qy    313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
          ||:|:|::|||  :|:   || ||  | |  |||| |:| || | |::|:|| ||| ::
Db    298 KSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy    373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
          ||:||| |  : |:|||| :||| :|||: :| ::
Db    358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


RESULT   5
Q9D6X0
ID   Q9D6X0      PRELIMINARY;      PRT;   395 AA.
AC   Q9D6X0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2310051B21RIK PROTEIN.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK009875; BAB26556.1; -.
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44665 MW;  40CA6B67859A8C5B CRC64;


  Query Match             50.5%;  Score 1134.5;  DB 11;  Length 395;
  Best Local Similarity   53.5%;  Pred. No. 1.8e-92;
  Matches  213;  Conservative   70;  Mismatches  108;  Indels    7;  Gaps    3;

Qy     17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
          ||||:: : :     : :        :||| ||:|::| : ||| ||||| ||||||| |
Db      1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy     76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
```

```
          ||| |    :  | |||  |||||:  |:|||:|||||| |||||||:|||:||||||
Db     61 YRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy    136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
          ||||:   | |  ||||||:||||::|||| |:||:|||||||:|||:|||||:||||
Db    121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy    196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
          ||| | ||:||| ::|||||:|||| :|| |  |:|  ::| :|| | |:   |  :||
Db    181 FSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPKFLLKVIFGNKMFMPHNYLDQFL 240

Qy    253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
          .|   | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db    241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRFDVYLGHNPAGTSTQDLFHWAQLA 297

Qy    313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
          ||:|:|::|||  :|:   || ||  | |  |||| |:| || | |::|:|| ||| ::
Db    298 KSGKLQAYNWGSPLRNMLHYNQKTPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy    373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
          ||:||| |  : |:|||| :|||  :||||: :| ::
Db    358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


RESULT   6
Q9D6Q6
ID   Q9D6Q6      PRELIMINARY;      PRT;   395 AA.
AC   Q9D6Q6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ADULT MALE TONGUE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE   CLONE:2310067K20, FULL INSERT SEQUENCE.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK010093; BAB26697.1; -.
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44623 MW;  245055F5E7FF07C9 CRC64;
```

```
  Query Match             50.5%;  Score 1134.5;  DB 11;  Length 395;
  Best Local Similarity   53.8%;  Pred. No. 1.8e-92;
  Matches  214;  Conservative   69;  Mismatches  108;  Indels    7;  Gaps    3;

Qy     17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
          ||||:: : :      : :       :||| ||:|::| : ||| ||||| |||||||
Db      1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy     76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
          ||| |    :  | |||  ||||||  |:|||:||||||| |||||||:|||:||||||
Db     61 YRIPYGKKNSENIGKRPVAYLQHGLVASATNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy    136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
          ||||:   | |  ||||||:||||::|||| |:|||:|||||||:|||:|||||:||||
Db    121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEK1HYVGHSQGTTIGFIA 180

Qy    196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
          ||| | ||:||| ::|||||:|||| :|| |  |:|  ::| :|| | |:   |  :||
Db    181 FSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPKFLLKVIFGNKMFMPHNYLDQFL 240

Qy    253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
          |  | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db    241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRFDVYLGHNPAGTSTQDLFHWAQLA 297

Qy    313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
          ||:|:|::|||  :|:   || ||  | |  |||| |:| || | |::|:|| ||| ::
Db    298 KSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy    373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
          ||:||| |  : |:|||| :|||  :||||: :| ::
Db    358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


RESULT   7
Q9D6P3
ID   Q9D6P3      PRELIMINARY;      PRT;   395 AA.
AC   Q9D6P3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ADULT MALE TONGUE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE   CLONE:2310069P19, FULL INSERT SEQUENCE.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK010139; BAB26725.1; -.
```

```
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44651 MW;  E48A73CCCBFD359D CRC64;


  Query Match           50.5%;  Score 1134.5;  DB 11;  Length 395;
  Best Local Similarity   53.8%;  Pred. No. 1.8e-92;
  Matches  214;  Conservative   69;  Mismatches  108;  Indels    7;  Gaps    3;

Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||||:: : :      : :       :||| ||:|::| : ||| ||||| ||||||| |
Db       1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |    :  | |||  |||||:. |:|||:||||||| |||||||:|||:||||||
Db      61 YRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           ||||:   | |  ||||||:||||::|||| |:||:|||||||||:|||:|||||:||||
Db     121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
           ||| | ||:||| ::|||||||||: :||  |  |:|  ::| :|| | |:   |  :||
Db     181 FSTNPALAKKIKRFYALAPIATVKYTESPFKKISLIPKFLLKVIFGNKMFMPHNYLDQFL 240

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
             |  | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db     241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRFDVYLGHNPAGTSTQDLFHWAQLA 297

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
           ||:|:|::|||  :|:   || ||  | |  |||| |:| || | |::|:|| ||| ::
Db     298 KSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ||:||| |  : |:|||| :|||  :|||: :| ::
Db     358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


RESULT   8
Q9CPP8
ID   Q9CPP8      PRELIMINARY;      PRT;   395 AA.
AC   Q9CPP8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2310051B21RIK PROTEIN.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK010116; BAB26711.1; -.
DR   EMBL; AK009300; BAB26201.1; -.
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44579 MW;  D2296865EA671D34 CRC64;


  Query Match           50.4%;  Score 1133.5;  DB 11;  Length 395;
  Best Local Similarity   53.5%;  Pred. No. 2.2e-92;
  Matches  213;  Conservative   70;  Mismatches  108;  Indels    7;  Gaps    3;

Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||||:: : :      : :       :||| ||:|::| : ||| ||||| ||||||| |
Db       1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |    :  | |||  |||||:  |:|||:||||||| |||||||:|||:||||||
Db      61 YRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           ||||:   | |  ||||||:||||::|||| |:||:|||||||||:|||:|||||:||||
Db     121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
           ||| | ||:||| ::|||||||||: :||  |  |:|  ::| :|| | |:   |  :||
Db     181 FSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPKFLLKVIFGNKMFMPHNYLDQFL 240

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
             |  | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db     241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRFGVYLGHNPAGTSTQDLFHWAQLA 297

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
           ||:|:|::|||  :|:   || ||  | |  |||| |:| || | |::|:|| ||| ::
Db     298 KSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ||:||| |  : |:|||| :|||  :|||: :| ::
Db     358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


RESULT   9
Q9CPP7
ID   Q9CPP7      PRELIMINARY;      PRT;   395 AA.
AC   Q9CPP7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2310051B21RIK PROTEIN.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
```

```
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK010236; BAB26787.1; -.
DR   EMBL; AK009413; BAB26272.1; -.
DR   EMBL; AK009428; BAB26280.1; -.
DR   EMBL; AK009459; BAB26300.1; -.
DR   EMBL; AK009473; BAB26312.1; -.
DR   EMBL; AK009474; BAB26313.1; -.
DR   EMBL; AK009479; BAB26316.1; -.
DR   EMBL; AK009523; BAB26338.1; -.
DR   EMBL; AK009525; BAB26339.1; -.
DR   EMBL; AK009546; BAB26352.1; -.
DR   EMBL; AK009571; BAB26368.1; -.
DR   EMBL; AK009573; BAB26370.1; -.
DR   EMBL; AK009729; BAB26466.1; -.
DR   EMBL; AK009773; BAB26495.1; -.
DR   EMBL; AK010019; BAB26647.1; -.
DR   EMBL; AK010035; BAB26656.1; -.
DR   EMBL; AK010058; BAB26673.1; -.
DR   EMBL; AK010061; BAB26675.1; -.
DR   EMBL; AK010124; BAB26715.1; -.
DR   EMBL; AK010125; BAB26716.1; -.
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44637 MW;  D3F96B65EA671E34 CRC64;

  Query Match             50.4%;  Score 1133.5;  DB 11;  Length 395;
  Best Local Similarity   53.5%;  Pred. No. 2.2e-92;
  Matches  213;  Conservative   70;  Mismatches  108;  Indels    7;  Gaps    3;

Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||||:: : :      : :       :||| ||:|::| : ||| ||||| ||||||| |
Db       1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |    :  | |||  ||||:  |:|||:|||||| |||||||:|||:||||||
Db      61 YRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           ||||:   | |  ||||||:||||::|||| |:||:||||||||:|||:|||||:||||
Db     121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
           ||| | ||:||| ::||||:||||: :||  |  |:|  ::| :|| | |:   |  :||
Db     181 FSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPKFLLKVIFGNKMFMPHNYLDQFL 240

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
           |   | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db     241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRFDVYLGHNPAGTSTQDLFHWAQLA 297

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
           ||:|:|::|||  :|:   || ||  | |  |||| |:| || | |::|:|| ||| ::
Db     298 KSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ||:||| |  : |:|||| :|||  :||||: :| ::
Db     358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


RESULT  10
Q9D6L9
ID   Q9D6L9      PRELIMINARY;      PRT;   395 AA.
AC   Q9D6L9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ADULT MALE TONGUE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE   CLONE:2310076L13, FULL INSERT SEQUENCE.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK010203; BAB26766.1; -.
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44671 MW;  E8936162510AA55C CRC64;

  Query Match             50.4%;  Score 1131.5;  DB 11;  Length 395;
  Best Local Similarity   53.5%;  Pred. No. 3.3e-92;
  Matches  213;  Conservative   69;  Mismatches  109;  Indels    7;  Gaps    3;

Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||||:: : :      : :       :||| ||:|::| : ||| ||||| ||||||| |
Db       1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |    :  | |||  ||||:  |:|||:|||||| |||||||:|||:||||||
Db      61 YRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           ||||:   | |  ||||||:||||::|||| |:||:||||||||:|||:|||||:||||
Db     121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
           ||| | ||:||| ::||||:||||: :||  |  |:|  ::| :|| | |:   |  :||
Db     181 FSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPKFFLKVIFGNKMFMPHNYLDQFL 240
```

```
Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
           |   | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db     241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRFDVYLGHNPAGTSTQDLFHWAQLA 297

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
           ||:|:|::|||  :|:   || ||  | |  |||| |:| || | |::|:|| ||| ::
Db     298 KSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ||:||| |  : |:|||| :|||  :||||: :| ::
Db     358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


RESULT  11
Q9D6T5
ID   Q9D6T5      PRELIMINARY;      PRT;   395 AA.
AC   Q9D6T5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ADULT MALE TONGUE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE   CLONE:2310061A13, FULL INSERT SEQUENCE.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK009990; BAB26629.1; -.
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44623 MW;  D3F96DB83161C3EF CRC64;


  Query Match             50.3%;  Score 1129.5;  DB 11;  Length 395;
  Best Local Similarity   53.3%;  Pred. No. 4.9e-92;
  Matches  212;  Conservative   71;  Mismatches  108;  Indels    7;  Gaps    3;

Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||||:: : :      : :       :||| ||:|::| : ||| ||||| ||||||| |
Db       1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |    :  | |||  |||||:  |:|||:||||||| |||||||:|||:||||||
Db      61 YRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           ||||:   | |  |||||:||||::|||| |:||:|||||||||:|||:|||||:||||
Db     121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
           ||| | ||:||| ::|||:|||: :||  |  |:|  ::| :|| | |:   |  :||
Db     181 FSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPKFLLKVIFGNKMFMPHNYLDQFL 240

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
           |   | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db     241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRFDVYLGHNPAGTSTQDLFHWAQLA 297

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
           ||:|:|::|||  :|:   || :|  | |  |||| |:| || | |::|:|| ||| ::
Db     298 KSGKLQAYNWGSPLQNMLHYNQKSPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ||:||| |  : |:|||| :|||  :||||: :| ::
Db     358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


RESULT  12
Q9D767
ID   Q9D767      PRELIMINARY;      PRT;   395 AA.
AC   Q9D767;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2310051B21RIK PROTEIN.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK009537; BAB26346.1; -.
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44607 MW;  E43317C2254FA8FB CRC64;


  Query Match             50.2%;  Score 1128.5;  DB 11;  Length 395;
  Best Local Similarity   53.3%;  Pred. No. 6.1e-92;
```

```
  Matches  212;  Conservative   70;  Mismatches  109;  Indels    7;  Gaps    3;

Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||||:: : :      : :        :||| ||:|::| : ||| ||||| ||||||| |
Db       1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |    :  | |||  ||||:  |:|||:|||||| |||||||:|||:||||||
Db      61 YRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           ||||:   | |  ||||||:||||::|||| |:||:|||||||:|||:|||| :||||
Db     121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEKIHYVGHSQGTAIGFIA 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
           ||| | ||:||| ::||||:|||| :||  |  |:|  ::| :|| | |:   |  :||
Db     181 FSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPKFLLKVIFGNKMFMPHNYLDQFL 240

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
           |   | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db     241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRFDVYLGHNPAGTSTQDLFHWAQLA 297

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
           ||:|:|::|||  :|:   || ||  | |  |||| |:| || | |::|:|| ||| ::
Db     298 KSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ||:||| |  : |:|||| :|||  :||||: :| ::
Db     358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


RESULT  13
Q9D7C5
ID   Q9D7C5      PRELIMINARY;      PRT;   395 AA.
AC   Q9D7C5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2310051B21RIK PROTEIN.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK009359; BAB26240.1; -.
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44748 MW;  3ED43992458DE058 CRC64;


  Query Match             50.2%;  Score 1127.5;  DB 11;  Length 395;
  Best Local Similarity   53.3%;  Pred. No. 7.4e-92;
  Matches  212;  Conservative   69;  Mismatches  110;  Indels    7;  Gaps    3;

Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||||:: : :      : :        :||| ||:|::| : ||| ||||| ||||||| |
Db       1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |    :  | |||  ||||:  |:|||:|||||| |||||||:|||:||||||
Db      61 YRIPYGKKNSENIGKRPVAYLQHGLIASAKNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           ||||:   | |  ||||||:||||::|||| |:||:|||||||:|||:|||| :||||
Db     121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
           ||| | ||:||| ::||||:|||| :||  |  |:|  ::| :|| | |:   |  :||
Db     181 FSTNPALAKKIKRFYALAPVATVKYTESPFKKIHFIPKFLLKVIFGNKMFMPHNYLDQFL 240

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
           |   | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db     241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRFDVYLGHNPAGTSTQDLFHWAQLA 297

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
           ||:|:|::|||  :|:   || ||  | |  |||| |:| || | |::|:|| ||| ::
Db     298 KSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ||:||| |  : |:|||| :|||  :||||: :| ::
Db     358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


RESULT  14
Q9D766
ID   Q9D766      PRELIMINARY;      PRT;   395 AA.
AC   Q9D766;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2310051B21RIK PROTEIN.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
```

```
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK009544; BAB26350.1; -.
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44648 MW;  CC69875653AA7A74 CRC64;


  Query Match             50.1%;  Score 1126.5;  DB 11;  Length 395;
  Best Local Similarity   53.3%;  Pred. No. 9.1e-92;
  Matches  212;  Conservative   70;  Mismatches  109;  Indels    7;  Gaps    3;

Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||||:: : :      : :       :||| ||:|::| : ||| ||||| ||||||| |
Db       1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |    :  | |||  | |||:  |:|||:|||||| |||||||:|||:||||||
Db      61 YRIPYGKKNSENIGKRPVAYLLHGLIASATNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           ||||:   | |  ||||||:||||::|||| |:||:|||||||||:|||:||||||:||||
Db     121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
           ||| | ||:||| ::||||:||||: :||  |  |:| :::| :|| | |:   |  :||
Db     181 FSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPKFLLKVIFGNKMFMPYNYLDQFL 240

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
           |   | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db     241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRFDVYLGHNPAGTSTQDLFHWAQLA 297

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
           ||:|:|::|||  :|:   || ||  | |  |||| |:| || | |::|:|| ||| ::
Db     298 KSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ||:||| |  : |:|||| :|||  :||||: :| ::
Db     358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


RESULT  15
Q9D6L1
ID   Q9D6L1      PRELIMINARY;      PRT;   395 AA.
AC   Q9D6L1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ADULT MALE TONGUE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE   CLONE:2310079O20, FULL INSERT SEQUENCE.
GN   2310051B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21065609; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR   EMBL; AK010231; BAB26784.1; -.
DR   MGD; MGI:1914967; 2310051B21Rik.
DR   InterPro; IPR000073; Abhydrolase.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
SQ   SEQUENCE   395 AA;  44659 MW;  13E0BB95FC5A40E0 CRC64;


  Query Match             50.1%;  Score 1126.5;  DB 11;  Length 395;
  Best Local Similarity   53.5%;  Pred. No. 9.1e-92;
  Matches  213;  Conservative   69;  Mismatches  109;  Indels    7;  Gaps    3;

Qy      17 MWLLILVAYMFQ-RNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDGYILSV 75
           ||||:: : :      : :       :||| ||:|::| : ||| ||||| ||||||| |
Db       1 MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYPSEEYEVVTEDGYILGV 60

Qy      76 NRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGN 135
           ||| |    :  | |||  |||||:  |:|||:|||||| |||||||:|||:||||||
Db      61 YRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLAFILADAGYDVWLGNSRGN 120

Qy     136 AWSRKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIA 195
           ||||:   | |  ||||||:||||::|||| |:||:|||||||||:|||:||||||:||||
Db     121 TWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKTGQEKIHYVGHSQGTTIGFIA 180

Qy     196 FSTMPELAQKIKMYFALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFL---YQTRFL 252
           ||| | ||:||| ::||||:||||: :||  |  |:|  ::| :|| | |:   |  :||
Db     181 FSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIHKFLLKVIFGNKMFMPHNYLDQFL 240

Qy     253 RQLVIYLCGQVILDQICSNIMLLLGGFNTNNMNMSRASVYAAHTLAGTSVQNILHWSQAV 312
           |   | : :|| :||| : :  ||:  |:|:||  ||  |  |||| |:: ||:|
Db     241 GTEV---CSRELLDLLCSNALFIFCGFDKKNLNVSRFDVYLGHNPAGTSTQDLFHWAQLA 297

Qy     313 NSGELRAFDWGSETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSNPEDVKMLLSEV 372
           ||:|:|::|||  :||   || ||  | |  |||| |:| || | |::|:|| ||| ::
Db     298 KSGKLQAYNWGSPLQNLLHYNQKTPPYYDVSAMTVPIAVWNGGHDILADPQDVAMLLPKL 357

Qy     373 TNLIYHKNIPEWAHVDFIWGLDAPHRMYNEIIHLMQQE 410
           ||:||| |  : |:|||| :|||  :||||: :| ::
Db     358 PNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMMAED 395


Search completed: January 30, 2002, 23:09:14
Job time: 550 sec
```

Jiang, L.
09/578063
Seq. IDs 46 & 47

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 30, 2002, 20:24:53 ; Search time 1858.36 Seconds
(without alignments)
11265.267 Million cell updates/sec

Title: US-09-578-063-46
Perfect score: 1269
Sequence: 1 atgttggaaaccttgtcaag.....gacggtgtgaggccgtattg 1269

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.		Score	Query Match	Length	DB	ID	Description
1		527.8	41.6	2481	9	HSLYACLY	Z31690 H.sapiens (
2		527.8	41.6	2481	9	HSU08464	U08464 Human lysos
3		527.8	41.6	2626	9	HSLAL	X76488 H.sapiens m
4		526.2	41.5	2493	9	HUMLIPCHL	M74775 Human lysos
5		524.6	41.3	2586	9	BC012287	BC012287 Homo sapi
6		489.2	38.6	1378	6	A26689	A26689 Precursor o
7	c	489.2	38.6	1378	6	A26690	A26690 Precursor o
8		470	37.0	1365	9	HSGLR	X05997 Human mRNA
9		470	37.0	1367	6	A01046	A01046 H.sapiens m
10		470	37.0	1367	6	A12714	A12714 Pregastric
11		467.6	36.8	1336	6	A01157	A01157 R.norvegicu
12		467.6	36.8	1355	10	RNLIP	X02309 Rat mRNA fo
13		461	36.3	2358	10	MMLYACLY	Z31689 M.musculus
14		459.8	36.2	3144	10	S81497	S81497 lysosomal a
15		458	36.1	1417	4	BOVMPE	L26319 Bos taurus
16		444.2	35.0	1137	6	AR039022	AR039022 Sequence
17		444.2	35.0	1137	6	AR092633	AR092633 Sequence
18		444.2	35.0	1140	6	AR039023	AR039023 Sequence
19		444.2	35.0	1140	6	AR092634	AR092634 Sequence
20		444.2	35.0	1146	6	AR039024	AR039024 Sequence
21		444.2	35.0	1146	6	AR092635	AR092635 Sequence
22		444.2	35.0	1198	6	A57760	A57760 Sequence 5
23		444.2	35.0	1528	6	A57756	A57756 Sequence 1
24		444.2	35.0	1528	6	AR039021	AR039021 Sequence
25		444.2	35.0	1528	6	AR092632	AR092632 Sequence
26		444.2	35.0	1651	4	CFLIPASE	Y13899 Canis famil
27		441	34.8	1137	6	A39301	A39301 Sequence 2
28		441	34.8	1140	6	A39303	A39303 Sequence 4
29		441	34.8	1146	6	A39305	A39305 Sequence 6
30		441	34.8	1528	6	A39300	A39300 Sequence 1
31		388.8	30.6	1048	6	A57758	A57758 Sequence 3
32		286	22.5	699	6	AX068249	AX068249 Sequence
33		286	22.5	1308	6	AX068259	AX068259 Sequence
34		269	21.2	43456	9	AL353113	AL353113 Human DNA
35	c	269	21.2	76653	2	AC015506	AC015506 Homo sapi
36	c	269	21.2	158195	2	AC015461	AC015461 Homo sapi
37		204.4	16.1	148720	2	AC019353	AC019353 Homo sapi
38		204.4	16.1	172834	2	AC011277	AC011277 Homo sapi
39		151.2	11.9	76653	2	AC015506	AC015506 Homo sapi
40		147	11.6	161826	9	AL358532	AL358532 Human DNA
41		140.6	11.1	401	9	HSLIPCHL04	U04288 Human lysos
42		140.6	11.1	1851	9	HSLIPA4	X75491 H.sapiens L
43	c	140.6	11.1	169404	2	AL353751	AL353751 Homo sapi
44	c	140.6	11.1	188147	2	AL513533	AL513533 Homo sapi
45		127.6	10.1	40114	3	CELZK6	AF067942 Caenorhab

ALIGNMENTS

RESULT 1
HSLYACLY
LOCUS HSLYACLY 2481 bp mRNA PRI 17-FEB-1997
DEFINITION H.sapiens (HepG2) LAL mRNA for lysosomal acid lipase.
ACCESSION Z31690
VERSION Z31690.1 GI:506430
KEYWORDS LAL; lipase; lysosomal acid lipase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2481)
AUTHORS Du,H. and Gregory,G.A.
TITLE Structural Conservation of Putative Functional Motifs between Mouse and Human Lysosomal Acid Lipase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2481)
AUTHORS Du,H.
TITLE Direct Submission
JOURNAL Submitted (05-APR-1994) Hong Du, Division of Human Genetics,

```
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH, 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1..2481
                     /organism="Homo sapiens"
                     /strain="HepG2"
                     /db_xref="taxon:9606"
                     /clone="clone hLAL"
                     /tissue_type="liver"
                     /cell_type="hepatoma"
                     /clone_lib="human liver HepG2 cDNA library/lambda ZAP"
     sig_peptide     34..96
                     /gene="LAL"
     gene            34..1233
                     /gene="LAL"
     CDS             34..1233
                     /gene="LAL"
                     /citation=[1]
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="CAA83495.1"
                     /db_xref="GI:506431"
                     /db_xref="SPTREMBL:Q16529"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     variation       79
                     /gene="LAL"
                     /citation=[1]
                     /replace="c"
     mat_peptide     97..1230
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     variation       100
                     /gene="LAL"
                     /citation=[1]
                     /replace="g"
     polyA_signal    2459..2464
     polyA_site      2480
BASE COUNT      692 a    477 c    513 g    799 t
ORIGIN

  Query Match            41.6%;  Score 527.8;  DB 9;  Length 2481;
  Best Local Similarity  65.4%;  Pred. No. 7.3e-147;
  Matches  791;  Conservative    0;  Mismatches  412;  Indels    6;  Gaps    1;

Qy       39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
Db       30 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTC 89

Qy       99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152
Db       90 TGAGGGGTCTAGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 149

Qy      153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212
Db      150 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 209

Qy      213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272
Db      210 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 269

Qy      273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332
Db      270 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 329

Qy      333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392
Db      330 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 389

Qy      393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgaatt 452
Db      390 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 449

Qy      453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512
Db      450 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 509

Qy      513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572
Db      510 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 569

Qy      573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632
Db      570 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 629

Qy      633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaattttttgttgct 692
Db      630 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 689

Qy      693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752
Db      690 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 749

Qy      753 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 812
Db      750 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 809

Qy      813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872
Db      810 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 869

Qy      873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932
Db      870 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 929

Qy      933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992
Db      930 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 989

Qy      993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052
Db      990 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1049

Qy     1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112
Db     1050 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1109

Qy     1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172
Db     1110 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1169

Qy     1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232
Db     1170 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1229

Qy     1233 gaccaacct 1241
Db     1230 GTGAAAGCT 1238

RESULT    2
HSU08464
LOCUS       HSU08464     2481 bp    mRNA         PRI       23-JUN-1994
DEFINITION  Human lysosomal acid lipase mRNA, complete cds.
ACCESSION   U08464
VERSION     U08464.1  GI:505052
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2481)
  AUTHORS   Du,H. and Gregory,G.A.
  TITLE     Structural Conservation of Putative Functional Motifs between Mouse
            and Human Lysosomal Acid Lipase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2481)
  AUTHORS   Du,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1994) Hong Du, Division of Human Gentics,
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2481
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="clone hLAL"
                     /clone_lib="HepG2 cDNA library/Lambda ZAP"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /tissue_type="liver"
     5'UTR           1. .33
     sig_peptide     34. .96
     CDS             34. .1233
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="AAB60328.1"
                     /db_xref="GI:505053"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     variation       79
                     /replace="c"
     mat_peptide     97. .1230
                     /product="lysosomal acid lipase"
     variation       100
                     /replace="g"
     3'UTR           1234. .2480
     polyA_signal    2459. .2464
     polyA_site      2480
BASE COUNT      692 a    477 c    513 g    799 t
ORIGIN
```

Query Match 41.6%; Score 527.8; DB 9; Length 2481;
Best Local Similarity 65.4%; Pred. No. 7.3e-147;
Matches 791; Conservative 0; Mismatches 412; Indels 6; Gaps 1;

```
Qy     39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
Db     30 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTC 89

Qy     99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152
Db     90 TGAGGGGTCTAGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 149

Qy    153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212
Db    150 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 209

Qy    213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272
Db    210 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 269

Qy    273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332
Db    270 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 329

Qy    333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392
Db    330 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 389

Qy    393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452
Db    390 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 449

Qy    453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512
Db    450 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 509

Qy    513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572
Db    510 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 569

Qy    573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632
Db    570 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 629

Qy    633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 692
Db    630 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 689

Qy    693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752
Db    690 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 749

Qy    753 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 812
Db    750 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 809

Qy    813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872
Db    810 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 869

Qy    873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932
Db    870 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 929

Qy    933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992
Db    930 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 989

Qy    993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052
Db    990 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1049

Qy   1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112
Db   1050 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1109

Qy   1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172
Db   1110 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1169

Qy   1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232
Db   1170 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1229

Qy   1233 gaccaacct 1241
Db   1230 GTGAAAGCT 1238
```

```
RESULT   3
HSLAL
LOCUS       HSLAL        2626 bp    mRNA            PRI       25-FEB-1994
DEFINITION  H.sapiens mRNA for lysosomal acid lipase.
ACCESSION   X76488
VERSION     X76488.1  GI:434305
KEYWORDS    lysosomal acid lipase.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2626)
  AUTHORS   Ameis,D., Merkel,M., Eckerskorn,C. and Greten,H.
  TITLE     Purification, characterization and molecular cloning of human
            hepatic lysosomal acid lipase
  JOURNAL   Eur. J. Biochem. 219 (3), 905-914 (1994)
  MEDLINE   94155897
REFERENCE   2  (bases 1 to 2626)
  AUTHORS   Ameis,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-1993) D. Ameis, Medical Department, University
            Hospital Eppendorf, Martinistrasse 52, 20246 Hamburg, FRG
FEATURES             Location/Qualifiers
     source          1. .2626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /tissue_type="liver"
     gene            146. .1345
                     /gene="LAL/LIPA"
     CDS             146. .1345
                     /gene="LAL/LIPA"
                     /EC_number="3.1.1.13"
                     /note="sterol esterase"
                     /codon_start=1
                     /evidence=experimental
                     /product="lysosomal acid lipase"
                     /protein_id="CAA54026.1"
                     /db_xref="GI:434306"
                     /db_xref="SWISS-PROT:P38571"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSGGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     sig_peptide     146. .226
                     /gene="LAL/LIPA"
                     /evidence=experimental
     misc_feature    227. .373
                     /gene="LAL/LIPA"
                     /note="propeptide"
                     /evidence=experimental
     mat_peptide     374. .1342
                     /gene="LAL/LIPA"
                     /evidence=experimental
                     /product="lysosomal acid lipase"
BASE COUNT      741 a    527 c    547 g    811 t
ORIGIN
```

```
  Query Match           41.6%;  Score 527.8;  DB 9;  Length 2626;
  Best Local Similarity 65.4%;  Pred. No. 7.4e-147;
  Matches  791;  Conservative    0;  Mismatches  412;  Indels    6;  Gaps    1;

Qy      39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
           ||||||| ||||| || |  ||    ||||||  | | || | |   | |   || ||||
Db     142 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTC 201

Qy      99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152
           ||      |  |   | ||||||||| || ||| ||  ||||||| | |||||
Db     202 TGAGGGGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 261

Qy     153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212
           ||| |||    |    || |  ||  ||||||||||   ||| |  || ||||||||  ||
Db     262 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 321

Qy     213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272
           ||| || |   |||||| | |||||||  ||      | | | |    | | |||| |||
Db     322 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 381

Qy     273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332
           || || || || ||||| |||||| |  | | || | |||| |||||| |  | |||||
Db     382 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 441

Qy     333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392
           ||||||  |||||||||||||||| || ||||||||||||||||||||||||||||| ||
Db     442 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 501

Qy     393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452
           ||||||  |||||  |||||||||| |||||  |||||||||||   |    || ||||| ||
Db     502 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 561

Qy     453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512
           |||||||||||||||||||||||||||  | |||||| || ||    || ||||| ||
Db     562 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 621

Qy     513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572
           | || | ||||| ||||| ||| |  | |||||||| ||  |||| || ||||||||  ||
Db     622 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 681

Qy     573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632
           || ||||| ||||||||    || || |||||||| | |  || |||||||  ||||||
Db     682 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 741

Qy     633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 692
           | |  ||  | ||  | ||           | |||||  | |||||||         |
Db     742 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 801

Qy     693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752
           ||||||    | || ||||  || |||||  | ||||||||||    ||||     ||
Db     802 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 861

Qy     753 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 812
           | | |      || | ||   || ||||    || || || ||  |  || | ||| | ||
Db     862 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 921

Qy     813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872
           | ||   || |||||| |||||| ||    |  ||| | || |||    |||    |||
Db     922 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 981

Qy     873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932
           |||| |  | ||  ||| |||||||||  |||||||||| ||  ||||||||||||||
Db     982 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 1041

Qy     933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992
           || || |      |  | |  || |||||||||| ||     |||| |||      |
Db    1042 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 1101

Qy     993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052
           | ||| |||     ||||   | ||||  || |  || |||   || || || ||| |
Db    1102 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1161

Qy    1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112
           ||||  || ||||| |||||||||| || ||    | |||||  || ||  | ||  || |
Db    1162 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1221

Qy    1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172
           | | |||||| |  | | |||| ||| |||||  |||||||  ||  | || |||||  ||
Db    1222 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1281

Qy    1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232
           |||  |||||| |||    |  | || ||| |||| ||  |||| |||  | |   | |
Db    1282 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1341

Qy    1233 gaccaacct 1241
           |   || ||
Db    1342 GTGAAAGCT 1350

RESULT   4
```

```
HUMLIPCHL
LOCUS       HUMLIPCHL    2493 bp    mRNA            PRI       07-JAN-1995
DEFINITION  Human lysosomal acid lipase/cholesteryl esterase mRNA, complete
            cds.
ACCESSION   M74775
VERSION     M74775.1  GI:187151
KEYWORDS    lysosomal acid lipase/cholesteryl esterase.
SOURCE      Homo sapiens fibroblast cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2493)
  AUTHORS   Anderson,R.A. and Sando,G.N.
  TITLE     Cloning and expression of cDNA encoding human lysosomal acid
            lipase/cholesteryl ester hydrolase. Similarities to gastric and
            lingual lipases
  JOURNAL   J. Biol. Chem. 266 (33), 22479-22484 (1991)
  MEDLINE   92042192
FEATURES             Location/Qualifiers
     source          1. .2493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="fibroblast"
                     /map="10"
     mRNA            <1. .2493
                     /gene="LIPA"
                     /note="G00-120-153"
     gene            1. .2493
                     /gene="LIPA"
     CDS             41. .1240
                     /gene="LIPA"
                     /EC_number="3.1.1.13"
                     /note="alternate initiator ATG may be at position 47"
                     /codon_start=1
                     /db_xref="GDB:G00-120-153"
                     /product="lysosomal acid lipase/cholesteryl esterase"
                     /protein_id="AAA59519.1"
                     /db_xref="GI:187152"
                     /translation="MKMRFLGLVVCLVLWPLHSEGSGGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     sig_peptide     41. .103
                     /gene="LIPA"
                     /note="alternate signal peptide may begin at position 47;
                     as predicted from von Heijne, G. (1986) Nucl. Acids Res.
                     14,4683-4690; G00-120-153"
     mat_peptide     104. .1237
                     /gene="LIPA"
                     /EC_number="3.1.1.13"
                     /note="G00-120-153"
                     /product="lysosomal acid lipase/cholesteryl esterase"
     polyA_signal    2466. .2471
                     /gene="LIPA"
                     /note="G00-120-153"
     polyA_site      2493
                     /gene="LIPA"
                     /note="G00-120-153"
BASE COUNT      692 a    484 c    516 g    801 t
ORIGIN
```

Query Match 41.5%; Score 526.2; DB 9; Length 2493;
Best Local Similarity 65.3%; Pred. No. 2.2e-146;
Matches 790; Conservative 0; Mismatches 413; Indels 6; Gaps 1;

```
Qy     39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
Db     37 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGCCCCTGCATTC 96

Qy     99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152
Db     97 TGAGGGGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 156

Qy    153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212
Db    157 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 216

Qy    213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272
Db    217 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 276

Qy    273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332
Db    277 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 336

Qy    333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392
Db    337 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 396

Qy    393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452
Db    397 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 456

Qy    453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512
Db    457 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 516

Qy    513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572
Db    517 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 576

Qy    573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632
Db    577 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 636

Qy    633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 692
Db    637 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 696

Qy    693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752
Db    697 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 756

Qy    753 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 812
Db    757 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 816

Qy    813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872
Db    817 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 876

Qy    873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932
Db    877 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 936

Qy    933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992
Db    937 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 996

Qy    993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052
Db    997 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1056

Qy   1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112
Db   1057 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1116

Qy   1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172
Db   1117 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1176
```

```
Qy   1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232
          |||  ||||||| |||    |  | || ||| |||| ||  |||| |||  | |  |  |
Db   1177 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1236

Qy   1233 gaccaacct 1241
          |   || ||
Db   1237 GTGAAAGCT 1245


RESULT  5
BC012287
LOCUS       BC012287     2586 bp    mRNA            PRI       08-AUG-2001
DEFINITION  Homo sapiens, Similar to lipase A, lysosomal acid, cholesterol
            esterase (Wolman disease), clone MGC:5229 IMAGE:2900168, mRNA,
            complete cds.
ACCESSION   BC012287
VERSION     BC012287.1  GI:15126726
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2586)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 3 Row: c Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 434305.
FEATURES             Location/Qualifiers
     source          1. .2586
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:5229 IMAGE:2900168"
                     /tissue_type="Placenta, choriocarcinoma"
                     /clone_lib="NIH_MGC_10"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             101. .1300
                     /codon_start=1
                     /product="Similar to lipase A, lysosomal acid, cholesterol
                     esterase (Wolman disease)"
                     /protein_id="AAH12287.1"
                     /db_xref="GI:15126727"
                     /translation="MKMRFLGLVVCLVLWPLHSEGSGGKLTALDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
BASE COUNT      736 a    511 c    531 g    808 t
```

```
ORIGIN

  Query Match          41.3%;  Score 524.6;  DB 9;  Length 2586;
  Best Local Similarity  65.3%;  Pred. No. 6.7e-146;
  Matches  789;  Conservative    0;  Mismatches  414;  Indels    6;  Gaps    1;

Qy     39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
          ||||||| ||||| || |  ||    ||||||  | | || | |    |     || ||||
Db     97 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGCCCCTGCATTC 156

Qy     99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152
           |      |  |   | |||| |||| || ||| ||  |||||||  | |||||
Db    157 TGAGGGGTCTGGAGGGAAACTGACAGCTTTGGATCCTGAAACAAACATGAATGTGAGTGA 216

Qy    153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212
          ||| |||    |    || |  ||  ||||||||||   ||| |  || |||||||| ||
Db    217 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 276

Qy    213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272
          ||| || |   ||||| | |||||||  ||      | | | |    | | |||| |||
Db    277 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 336

Qy    273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332
          || || || || ||||| |||||| |  | | || | |||| |||||| |  | |||||
Db    337 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 396

Qy    333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392
          ||||||  ||||||||||||||||| || ||||||||||||||||||||||||||| ||
Db    397 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 456

Qy    393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452
          ||||||  |||||  |||||||||| |||||  |||||||||||   |    || ||||| ||
Db    457 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 516

Qy    453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512
          ||||||||||||||||||||||||||||  | ||||||  || ||    || |||||  ||
Db    517 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 576

Qy    513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572
          | || | |||||  |||||  ||| |  | ||||||| ||  |||| || |||||||  ||
Db    577 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 636

Qy    573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632
          || |||||  |||||||     || || |||||||| | |  || |||||||  ||||||
Db    637 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 696

Qy    633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 692
          | |  ||  | ||  | ||          |  |||||   | |||||||          |
Db    697 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 756

Qy    693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752
          ||||||    | || ||||  || |||||  | ||||||||||||    ||||     ||
Db    757 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 816

Qy    753 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 812
          | | |     || | ||   || ||||    || || || ||  |  || | ||| | ||
Db    817 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 876

Qy    813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872
          | ||   || |||||| |||||| ||    |  ||| | || |||    |||    |||
Db    877 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 936

Qy    873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932
          |||| |  | ||  ||| |||||||||| ||||||||||| ||  ||||||||||||||||
Db    937 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 996

Qy    933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992
          || || |     |  | |  || |||||||||| ||    |||| |||        |
Db    997 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 1056

Qy    993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052
          | ||| |||     ||||  |  ||||  || |  || |||   || || || ||| |
```

```
Db   1057 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1116

Qy   1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112
          |||| || ||||| |||||||||| || ||    | ||||| || ||  | ||  || |
Db   1117 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1176

Qy   1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172
          | | |||||| |  | | |||| |||  ||||| |||||||  ||  | || ||||| ||
Db   1177 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1236

Qy   1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232
          |||  ||||||| |||    |  | || ||| |||| ||  |||| ||| | |  |  |
Db   1237 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1296

Qy   1233 gaccaacct 1241
          |   || ||
Db   1297 GTGAAAGCT 1305


RESULT   6
A26689
LOCUS       A26689       1378 bp    DNA             PAT       05-APR-1995
DEFINITION  Precursor of rabit gastric lipase coding sequence.
ACCESSION   A26689
VERSION     A26689.1  GI:905029
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Benicourt,C., Blanchard,C. and Junien,J.L.
  TITLE     Recombinant gastric lipase from rabbit and pharmaceutical
            compositions
  JOURNAL   Patent: EP 0542629-A 9 19-MAY-1993;
            INSTITUT DE RECHERCHE JOUVEINAL
FEATURES             Location/Qualifiers
     source          1. .1378
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT      397 a    287 c    276 g    418 t
ORIGIN


  Query Match             38.6%;  Score 489.2;  DB 6;  Length 1378;
  Best Local Similarity   65.0%;  Pred. No. 2.5e-135;
  Matches  722;  Conservative    0;  Mismatches  388;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||     |||||||||||| | || ||     |    || || ||  ||||
Db    136 ACCCTGAAGTGAATATGAATATTAGTCAGATGATTTCCTACTGGGGATACCCAAGTGAAA 195

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          |||||||||| |  |||||||||| |||||||||   |||||||| ||||||   ||
Db    196 AATATGAAGTTGTGACTGAAGATGGTTATATCCTTGAGGTTAACAGAATTCCTTATGGGA 255

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  |   ||| | |||    || || || || ||  ||||||||||  |  |||
Db    256 AGAAAAATTCAGGGAACAGAGGCCAGAGACCCGTCGTATTTTTGCAGCATGGTTTGCTTG 315

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   ||    |||||||| |||||||||||||||| ||||||| ||| ||||||||||
Db    316 CATCAGCCTCAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTTATTCTGGCAG 375

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          |||||||| ||  |||||| |||||||||||||| |||||  |||||||  ||| | ||
Db    376 ATGCTGGTTATGGTGTGTGGCTGGGGAACAGCAGAGGAAATACCTGGTCCAGAAGAAACT 435

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
              ||   ||||   | ||| |||||||||| || | |||||| |||||||  |
Db    436 TATACTATTCACCAGACTCCGTTGAATTCTGGGCTTTTAGCTTTGATGAAATGGCTAAAT 495

Qy    485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
          ||||||||| ||    ||  |||| ||| |  || |||| || |||||||||  |  | |
```

```
Db    496 ATGACCTTCCAGCCACAATTGACTTCATTGTAAAGGAAACTGGACAGGAAAAGCTGCACT 555

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          |||| ||  |||| ||||||||||||||| || |||||||| ||||| || |  ||  |||
Db    556 ATGTTGGTCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCTACAAATCCCAAGC 615

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||| | | |||||||| | || ||| ||||| ||  | |||||||| ||| || |||
Db    616 TGGCTGAAAGAATCAAAACCTTTTATGCATTAGCTCCAGTCGCCACTGTGAAGTATACAA 675

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||||||  |  | |||| ||  |||  | ||    |  ||| |||||    | |||||
Db    676 AAAGCCTTGTAAACAAACTTAGGTTTATTCCTCCAACCATGTTCAAGATTATATTTGGTG 735

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||        || |     |||      |   ||||  | | |  | |||  ||
Db    736 ACAAAATATTCTACCCACACAATTTCTTTGATCAATTTCTTGCCACTCAAGTGTGTTCCC 795

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||  ||  ||   ||||| || |||  | | ||  |  |  |||| ||  ||| |
Db    796 GTGAGACACTGAATGTAATTTGCAGCAATGCCTTATTTATCATTTGTGGCTTTGACAGCG 855

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  |||||||||| ||      ||| || |   | || | ||  || ||||| || |
Db    856 CAAACTTGAACATGAGTCGCTTGGATGTGTACGTGTCACATAATCCGGCAGGAACTTCAG 915

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | ||||| || || ||||||| |||||| || || |||||  |  | |  || ||| | |
Db    916 TTCAAAACATGCTGCACTGGACCCAGGCTGTTAAATCTGGAAATTTTCAAGCTTTTAATT 975

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| |||    |  | ||| | |   | | |||||||| || |||      |||| |
Db    976 GGGGAAGTCCAGCTCAGAATGTAGTGCACTTCAATCAGCCCACACCTCCCTACTACAATG 1035

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | |  |  |||| || || |  ||| |||||  || ||  | |||||| |  |  | |
Db   1036 TGACCGCCATGAATGTGCCAATTGCAGTGTGGAGTGGTGGCAATGACTGGTTGGCTGACC 1095

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  |||| ||  |  |  |||| |  |  |  | || ||||| ||||| ||| | ||||
Db   1096 CCCAAGATGTTGACCTTTTGCTTCCAAAACTTTCTAATCTCATTTACCACAAGGAGATTC 1155

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcacgtatgtacaatg 1204
          | ||   |||| ||||||| |||||||   || ||||||||||     | |||||||
Db   1156 TTCCATATAATCACTTGGATTTTATCTGGGCAATGAATGCTCCTCAAGAAGTTTACAATG 1215

Qy   1205 aaatcatccatctgatgcagcaggaggaga 1234
          |||| ||   | |||||    | ||  | |
Db   1216 AAATTATTTCTATGATGGCAAAAGATAAAA 1245


RESULT   7
A26690/c
LOCUS       A26690       1378 bp    DNA             PAT       05-APR-1995
DEFINITION  Precursor of rabit gastric lipase (comp.).
ACCESSION   A26690
VERSION     A26690.1  GI:905030
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Benicourt,C., Blanchard,C. and Junien,J.L.
  TITLE     Recombinant gastric lipase from rabbit and pharmaceutical
            compositions
  JOURNAL   Patent: EP 0542629-A 10 19-MAY-1993;
            INSTITUT DE RECHERCHE JOUVEINAL
FEATURES             Location/Qualifiers
     source          1. .1378
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT      418 a    276 c    287 g    397 t
ORIGIN
```

```
  Query Match            38.6%;  Score 489.2;  DB 6;  Length 1378;
  Best Local Similarity  65.0%;  Pred. No. 2.5e-135;
  Matches  722;  Conservative    0;  Mismatches  388;  Indels    0;  Gaps    0;

Qy     125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
           |||| ||||     |||||||||||| | || ||      |    || || ||  ||||
Db    1243 ACCCTGAAGTGAATATGAATATTAGTCAGATGATTTCCTACTGGGGATACCCAAGTGAAA 1184

Qy     185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
           |||||||||| |  |||||||||| |||||||||   |||||||| ||||||   ||
Db    1183 AATATGAAGTTGTGACTGAAGATGGTTATATCCTTGAGGTTAACAGAATTCCTTATGGGA 1124

Qy     245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
             |  |   ||| | |||    || || || || ||  ||||||||||  |  |||
Db    1123 AGAAAAATTCAGGGAACAGAGGCCAGAGACCCGTCGTATTTTTGCAGCATGGTTTGCTTG 1064

Qy     305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
             |   ||    |||||||| |||||||||||||||||| |||||| ||| ||||||||||
Db    1063 CATCAGCCTCAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTTATTCTGGCAG 1004

Qy     365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
           |||||||| ||  |||||| |||||||||||||| |||||  |||||||  ||| | ||
Db    1003 ATGCTGGTTATGGTGTGTGGCTGGGGAACAGCAGAGGAAATACCTGGTCCAGAAGAAACT 944

Qy     425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
               ||   ||||    | ||| ||||||||||| || | |||||| |||||||  |
Db     943 TATACTATTCACCAGACTCCGTTGAATTCTGGGCTTTTAGCTTTGATGAAATGGCTAAAT 884

Qy     485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
           |||||||| ||    ||  |||| || |  || |||| || |||||||| |  | |
Db     883 ATGACCTTCCAGCCACAATTGACTTCATTGTAAAGGAAACTGGACAGGAAAAGCTGCACT 824

Qy     545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
           |||| ||  |||| ||||||||||||||| || |||||||| ||||| || |  ||  |||
Db     823 ATGTTGGTCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCTACAAATCCCAAGC 764

Qy     605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
           ||||| | | ||||||||  | || ||| ||||| ||  | ||||||| ||| || |||
Db     763 TGGCTGAAAGAATCAAAACCTTTTATGCATTAGCTCCAGTCGCCACTGTGAAGTATACAA 704

Qy     665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
           ||||||  |  | |||| ||  |||  | ||    |  ||| |||||    | |||||
Db     703 AAAGCCTTGTAAACAAACTTAGGTTTATTCCTCCAACCATGTTCAAGATTATATTTGGTG 644

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
           | |||  |||       || |     |||     |  ||||  | | |  | |||  ||
Db     643 ACAAAATATTCTACCCACACAATTTCTTTGATCAATTTCTTGCCACTCAAGTGTGTTCCC 584

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
             | ||  ||  ||   ||||| || |||  | | ||  |  |  |||| ||  ||| |
Db     583 GTGAGACACTGAATGTAATTTGCAGCAATGCCTTATTTATCATTTGTGGCTTTGACAGCG 524

Qy     845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
             ||  |||||||||| ||      ||| || |   | || | ||  ||  ||||| || |
Db     523 CAAACTTGAACATGAGTCGCTTGGATGTGTACGTGTCACATAATCCGGCAGGAACTTCAG 464

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
           | ||||| || || ||||||| |||||| || || |||||  |  | |  || ||| | |
Db     463 TTCAAAACATGCTGCACTGGACCCAGGCTGTTAAATCTGGAAATTTTCAAGCTTTTAATT 404

Qy     965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
           |||| |||     |  | ||| | |    | | |||||||||| || |||      ||||  |
Db     403 GGGGAAGTCCAGCTCAGAATGTAGTGCACTTCAATCAGCCCACACCTCCCTACTACAATG 344

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           | |  |  |||| ||  || | ||| |||||| || ||  | |||||| |  |  | |
Db     343 TGACCGCCATGAATGTGCCAATTGCAGTGTGGAGTGGTGGCAATGACTGGTTGGCTGACC 284

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           |  |||| ||  |  |  ||||  |  |  | || ||||| ||||| ||| | ||||
Db     283 CCCAAGATGTTGACCTTTTGCTTCCAAAACTTTCTAATCTCATTTACCACAAGGAGATTC 224
```

```
Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           |  ||    |||| ||||||| |||||||   || ||||||||||     | |||||||
Db     223 TTCCATATAATCACTTGGATTTTATCTGGGCAATGAATGCTCCTCAAGAAGTTTACAATG 164

Qy    1205 aaatcatccatctgatgcagcaggaggaga 1234
           |||| ||   | |||||    | ||  | |
Db     163 AAATTATTTCTATGATGGCAAAAGATAAAA 134


RESULT   8
HSGLR
LOCUS       HSGLR        1365 bp    mRNA            PRI       30-MAR-1995
DEFINITION  Human mRNA for gastric lipase.
ACCESSION   X05997
VERSION     X05997.1  GI:31771
KEYWORDS    gastric lipase; lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1365)
  AUTHORS   Bodmer,M.W., Angal,S., Yarranton,G.T., Harris,T.J., Lyons,A.,
            King,D.J., Pieroni,G., Riviere,C., Verger,R. and Lowe,P.A.
  TITLE     Molecular cloning of a human gastric lipase and expression of the
            enzyme in yeast
  JOURNAL   Biochim. Biophys. Acta 909 (3), 237-244 (1987)
  MEDLINE   87299724
COMMENT     Data kindly reviewed (10-DEC-1987) by LOWE P.A.
FEATURES             Location/Qualifiers
     source          1. .1365
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     sig_peptide     45. .101
                     /product="pot.signal peptide (AA -19 to -1)"
     CDS             45. .1241
                     /codon_start=1
                     /product="gastric lipase precursor"
                     /protein_id="CAA29413.1"
                     /db_xref="GI:758063"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MWLLLTMASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGY
                     PNEEYEVVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSL
                     AFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKK
                     TGQKQLHYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVP
                     QSLFKFIFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLD
                     VYLSHNPAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVP
                     IAVWNGGKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSM
                     ISEDKK"
     CDS             63. .1241
                     /codon_start=1
                     /product="gastric lipase precursor"
                     /protein_id="CAA29414.1"
                     /db_xref="GI:758064"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGYPNEEYE
                     VVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSLAFILAD
                     AGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKKTGQKQL
                     HYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVPQSLFKF
                     IFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLDVYLSHN
                     PAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVPIAVWNG
                     GKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSMISEDKK
                     "
     sig_peptide     63. .101
                     /product="pot.signal peptide (AA -13 to -1)"
     mat_peptide     102. .1238
                     /product="gastric lipase (AA 1-379)"
BASE COUNT      396 a    284 c    268 g    417 t
ORIGIN


  Query Match            37.0%;  Score 470;  DB 9;  Length 1365;
  Best Local Similarity  64.2%;  Pred. No. 1.4e-129;
```

```
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy     119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
           |||    ||| ||||     ||||| |||||| | || ||      ||   || || ||
Db     121 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGGGGATACCCAA 180

Qy     179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
           ||| ||||||||||| |  ||||||||||| ||||| |||    || || || ||||||
Db     181 ATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCTTGAAGTCAATAGAATTCCTT 240

Qy     239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
           ||       |  |   ||| |||||     || ||||| |||||  ||||||||||
Db     241 ATGGGAAGAAAAATTCAGGGAATACAGGCCAGAGACCTGTTGTGTTTTTGCAGCATGGTT 300

Qy     299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
           |  ||| |   || |  |||||||||||||||||||| ||||| ||||| | ||||||||
Db     301 TGCTTGCATCAGCCACAAACTGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTC 360

Qy     359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
           ||||||||||||||| ||| |||||| |||| |||||||| |||||| ||||| |  |||
Db     361 TGGCAGATGCTGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 420

Qy     419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
           | ||  |       ||   |||   || ||| ||||||||||||||| | |||||| ||||
Db     421 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 480

Qy     479 ctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaaga 538
           |||  | |||||||| ||    ||  |||| ||| |  ||||||| || ||| |  ||
Db     481 CTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAAGAAAACTGGACAGAAGCAGC 540

Qy     539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
           |  | |||||  ||| |||| |||||||||||||| || ||||||||  ||||||||||  |
Db     541 TACACTATGTTGGCCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCCACCAATC 600

Qy     599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
           |    |||||| | | ||||||||  |  | |||| |||| ||  | ||||||| |||
Db     601 CCAGCCTGGCTAAAAGAATCAAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGT 660

Qy     659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
           || |||||||||     | |||| ||   ||  | ||  |     |  |||||   | |
Db     661 ATACAAAAAGCCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 720

Qy     719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 778
           ||||  | |||  |||       || | |  |||     |  |||| || | |  | |
Db     721 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCTACTGAAGTGT 780

Qy     779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
           |   ||  | ||| ||  |||   |||| || |||  | |  ||  |  |  |||||||
Db     781 GCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTTATTTATAATTTGTGGATTTG 840

Qy     839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
           |||  || ||  | |||| ||| ||      ||| |||    | || | ||  || ||||
Db     841 ACAGTAAGAACTTTAACACGAGTCGCTTGGATGTGTATCTATCACATAATCCAGCAGGAA 900

Qy     899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
           | ||||| ||||| ||  | || |||| |||||| || || |||||  || |||  || |
Db     901 CTTCTGTTCAAAACATGTTCCATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTT 960

Qy     959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
           |||||||||  ||        | |||  |    | |  |||||  |    |||      |
Db     961 ATGACTGGGGAAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1020

Qy    1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
           |||  || | ||  |||| || || |  ||| |||||  || ||  |||||  | |
Db    1021 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAGGACCTGTTGG 1080

Qy    1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
           |  | ||  |||| ||      |  ||||  |  |  |  |||| || || ||||| |||
Db    1081 CTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCCCAATCTTATTTACCACAAGG 1140

Qy    1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
           | ||||||   |    |||| |||| || |||||||   |||||| |||||     | |
Db    1141 AGATTCCTTTTTACAATCACTTGGACTTTATCTGGGCAATGGATGCCCCTCAAGAAGTTT 1200

Qy    1199 acaatgaaatcatccatctgat 1220
           |||||| || |   | ||||
Db    1201 ACAATGACATTGTTTCTATGAT 1222


RESULT   9
A01046
LOCUS       A01046       1367 bp    mRNA            PAT       16-MAR-1993
DEFINITION  H.sapiens mRNA for human gastric lipase.
ACCESSION   A01046
VERSION     A01046.1  GI:344241
KEYWORDS    gastric lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1367)
  AUTHORS   .
  TITLE     POLYPEPTIDE AND POLYPEPTIDE COMPOSITION
  JOURNAL   Patent: WO 8601532-A 5 13-MAR-1986;
FEATURES             Location/Qualifiers
     source          1. .1367
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     sig_peptide     47. .103
     CDS             47. .1243
                     /note="gastric lipase"
                     /codon_start=1
                     /product="lipase"
                     /protein_id="CAA00125.1"
                     /db_xref="GI:344242"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MWLLLTMASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGY
                     PNEEYEVVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSL
                     AFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKK
                     TGQKQLHYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVP
                     QSLFKFIFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLD
                     VYLSHNPAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVP
                     IAVWNGGKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSM
                     ISEDKK"
     mat_peptide     104. .1240
                     /note="gastric lipase"
                     /product="lipase"
BASE COUNT      397 a    284 c    269 g    417 t
ORIGIN


  Query Match             37.0%;  Score 470;  DB 6;  Length 1367;
  Best Local Similarity   64.2%;  Pred. No. 1.4e-129;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy     119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
           |||    ||| ||||     ||||| |||||| | || ||      ||   || || ||
Db     123 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGGGGATACCCAA 182

Qy     179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
           ||| ||||||||||| |  ||||||||||| ||||| |||    || || || ||||||
Db     183 ATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCTTGAAGTCAATAGAATTCCTT 242

Qy     239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
           ||       |  |   ||| |||||     || ||||| |||||  ||||||||||
Db     243 ATGGGAAGAAAAATTCAGGGAATACAGGCCAGAGACCTGTTGTGTTTTTGCAGCATGGTT 302

Qy     299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
           |  ||| |   || |  |||||||||||||||||||| ||||| ||||| | ||||||||
Db     303 TGCTTGCATCAGCCACAAACTGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTC 362

Qy     359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
           ||||||||||||||| ||| |||||| |||| |||||||| |||||| ||||| |  |||
Db     363 TGGCAGATGCTGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 422

Qy     419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
```

```
Db    423 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 482

Qy    479 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 538
Db    483 CTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAAGAAAACTGGACAGAAGCAGC 542

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
Db    543 TACACTATGTTGGCCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCCACCAATC 602

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
Db    603 CCAGCCTGGCTAAAAGAATCAAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGT 662

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
Db    663 ATACAAAAAGCCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 722

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 778
Db    723 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCTACTGAAGTGT 782

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
Db    783 GCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTTATTTATAATTTGTGGATTTG 842

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
Db    843 ACAGTAAGAACTTTAACACGAGTCGCTTGGATGTGTATCTATCACATAATCCAGCAGGAA 902

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
Db    903 CTTCTGTTCAAAACATGTTCCATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTT 962

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
Db    963 ATGACTGGGGAAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1022

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
Db   1023 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAGGACCTGTTGG 1082

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
Db   1083 CTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCCCAATCTTATTTACCACAAGG 1142

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
Db   1143 AGATTCCTTTTTACAATCACTTGGACTTTATCTGGGCAATGGATGCCCCTCAAGAAGTTT 1202

Qy   1199 acaatgaaatcatccatctgat 1220
Db   1203 ACAATGACATTGTTTCTATGAT 1224
```

```
RESULT  10
A12714
LOCUS       A12714        1367 bp    DNA             PAT       09-AUG-1994
DEFINITION  Pregastric lipase.
ACCESSION   A12714
VERSION     A12714.1  GI:579560
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1367)
  AUTHORS   .
  TITLE     PRECURSOR POLYPEPTIDE, DNA SEQUENCE CODING THEREFOR, VECTORS, HOST
            ORGANISMS, AND PROCESSES INVOLVING SAME
  JOURNAL   Patent: WO 8603778-A 8 03-JUL-1986;
FEATURES             Location/Qualifiers
     source          1. .1367
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             47. .1243
                     /codon_start=1
                     /product="pregastric lipase"
                     /protein_id="CAA01053.1"
                     /db_xref="GI:579561"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MWLLLTMASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGY
                     PNEEYEVVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSL
                     AFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKK
                     TGQKQLHYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVP
                     QSLFKFIFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLD
                     VYLSHNPAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVP
                     IAVWNGGKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSM
                     ISEDKK"
BASE COUNT      397 a    284 c    269 g    417 t
ORIGIN
```

Query Match 37.0%; Score 470; DB 6; Length 1367;
Best Local Similarity 64.2%; Pred. No. 1.4e-129;
Matches 707; Conservative 0; Mismatches 395; Indels 0; Gaps 0;

```
Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
Db    123 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGGGGATACCCAA 182

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
Db    183 ATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCTTGAAGTCAATAGAATTCCTT 242

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
Db    243 ATGGGAAGAAAAATTCAGGGAATACAGGCCAGAGACCTGTTGTGTTTTTGCAGCATGGTT 302

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
Db    303 TGCTTGCATCAGCCACAAACTGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTC 362

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
Db    363 TGGCAGATGCTGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 422

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
Db    423 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 482

Qy    479 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 538
Db    483 CTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAAGAAAACTGGACAGAAGCAGC 542

Qy    539 tctattatgtcggctattcacagggcaccacatgggctttattgcattttccaccatgc 598
Db    543 TACACTATGTTGGCCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCCACCAATC 602

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
Db    603 CCAGCCTGGCTAAAAGAATCAAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGT 662

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
Db    663 ATACAAAAAGCCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 722

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 778
Db    723 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCTACTGAAGTGT 782

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
Db    783 GCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTTATTTATAATTTGTGGATTTG 842
```

```
Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
          |||  || ||  | |||| ||| ||      ||| |||    | || | ||  || ||||
Db    843 ACAGTAAGAACTTTAACACGAGTCGCTTGGATGTGTATCTATCACATAATCCAGCAGGAA 902

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
          | ||||| ||||| ||  | || |||| |||||| || || |||||  || |||  || |
Db    903 CTTCTGTTCAAAACATGTTCCATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTT 962

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
          |||||||||  ||        | |||  |    | |   |||||  |    |||      |
Db    963 ATGACTGGGGAAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1022

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
          |||  || | ||  |||| || || |  ||| ||||||  || ||  |||||  | |
Db   1023 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAGGACCTGTTGG 1082

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
          |  | ||  |||| ||     |  ||||  |  |  |  |||| || || ||||| |||
Db   1083 CTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCCCAATCTTATTTACCACAAGG 1142

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
          | ||||||   |    |||| |||| || |||||||   ||||||| |||||     | |
Db   1143 AGATTCCTTTTTACAATCACTTGGACTTTATCTGGGCAATGGATGCCCCTCAAGAAGTTT 1202

Qy   1199 acaatgaaatcatccatctgat 1220
          |||||||  ||  |   | ||||
Db   1203 ACAATGACATTGTTTCTATGAT 1224
```

```
RESULT  11
A01157
LOCUS       A01157       1336 bp    mRNA            PAT       29-JAN-1993
DEFINITION  R.norvegicus mRNA for prelingual lipase protein.
ACCESSION   A01157
VERSION     A01157.1  GI:14748
KEYWORDS    lipase.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1336)
  AUTHORS   Williamson,R.
  JOURNAL   Patent: WO 8500381-A 3 31-JAN-1985;
            CellTech Ltd
FEATURES             Location/Qualifiers
     source          1. .1336
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     CDS             7. .1194
                     /codon_start=1
                     /product="prelingual lipase"
                     /protein_id="CAA00136.1"
                     /db_xref="GI:14749"
                     /db_xref="SWISS-PROT:P04634"
                     /translation="MWLLLITSVISTFGGAHGLFGKLGPGNPEANMNISQMITYWGYP
                     CQEYEVVTEDGYILGVYRIPHGKNNSENIGKRPVVYLQHGLIASATNWIANLPNNSLA
                     FMLADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATINFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPTLAKKIKTFYALAPVATVKYTQSPLKKISFIPT
                     FLFKLMFGKKMFLPHTYFDDFLGTEVCSREVLDLLCSNTLFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSVQDFLHWAQLVRSGKFQAFNWGSPSQNMLHYNQKTPPEYDVSAMTVPV
                     AVWNGGNDILADPQDVAMLLPKLSNLLFHKEILAYNHLDFIWAMDAPQEVYNEMISMM
                     AED"
BASE COUNT      409 a    283 c    265 g    379 t
ORIGIN
```

```
  Query Match             36.8%;  Score 467.6;  DB 6;  Length 1336;
  Best Local Similarity   63.5%;  Pred. No. 7.6e-129;
  Matches  713;  Conservative    0;  Mismatches  409;  Indels    0;  Gaps    0;

Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
          |||   |||| ||||||   |||||||||||| | || ||     |    || ||||| |
Db     80 CTGGAAACCCTGAAGCAAATATGAATATTAGTCAGATGATAACTTACTGGGGATATCCAT 139

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
          || | |||||||||||| |  |||||||||||| || || ||    ||  |||| |||||||
Db    140 GTCAAGAATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTACAGAATTCCTC 199

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
            ||       |  || | ||  | |||    || ||||||||||   ||||||||||
Db    200 ATGGGAAGAATAATTCTGAAAATATAGGCAAGAGACCTGTGGTGTATTTGCAGCATGGTT 259

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
          |  | | |   || |  |||||||||| | || || || ||||| ||||||| ||| || |
Db    260 TGATCGCATCAGCCACAAACTGGATTGCAAATCTACCAAACAACAGCCTGGCCTTTATGC 319

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
          | ||||| ||||| | ||| |||||| ||||||||||  | |||||  | |||||  | |
Db    320 TAGCAGACGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGAAATACATGGTCCAGGA 379

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
          || |        |||   ||||  || ||| ||||||||||||||  | ||||||  ||||
Db    380 AAAATGTATACTACTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 439

Qy    479 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 538
          |||  | ||||||||| ||    |||||||| ||| | |||||||| || || ||||||
Db    440 CTAAATATGACCTTCCCGCCACAATAAACTTCATTGTACAGAAAACTGGACAAGAGAAGA 499

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
          |  | ||||| ||  |||| |||||||||| || || || ||||| ||||| ||||  |
Db    500 TACACTATGTTGGTCATTCTCAGGGCACCACTATTGGTTTCATTGCCTTTTCTACCAATC 559

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
          |    |||||  | |||||||| | || || ||| |||||  ||  | || || || |||
Db    560 CTACACTGGCCAAAAAAATCAAGACGTTTTATGCATTAGCTCCAGTTGCTACCGTGAAGT 619

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
          || || |||| ||| || || ||| || | ||      |  | |||||    |||
Db    620 ATACACAAAGTCCCTTGAAAAAGATTTCATTTATTCCTACATTTCTTTTCAAGCTTATGT 679

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 778
          ||||||| |||   || |||   || |||   |||          |||| ||   |  | |
Db    680 TTGGCAAGAAAATGTTCCTGCCCCACACCTACTTTGATGACTTTCTTGGTACCGAAGTGT 739

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
          |    | || | |||| ||||   | || || || |   | ||  |  |  |||||||
Db    740 GCTCACGGGAGGTTCTAGATCTTCTCTGCAGCAACACTTTATTCATCTTCTGTGGATTTG 799

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
          |||  || ||  | ||  |||| ||      ||| |||    | || | ||  || ||||
Db    800 ACAAGAAAAACTTAAATGTGAGTCGTTTTGATGTGTATCTAGGGCATAATCCAGCAGGGA 859

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
          ||||||| ||| |  |||| ||||||   |||   || |  |||||  || | |  || |
Db    860 CATCTGTTCAAGACTTTCTCCACTGGGCACAGCTTGTTAGATCTGGGAAATTTCAAGCCT 919

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
          |  |||||||  ||      ||  | ||  ||  | | | |||  |||  |||  |||      |
Db    920 TCAACTGGGGAAGCCCATCCCAGAACATGTTACACTACAACCAGAAAACGCCTCCTGAAT 979

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
          |    ||   ||  ||||| || ||    ||| ||||||  || ||  | |||   ||
Db    980 ATGATGTGTCAGCCATGACTGTGCCAGTTGCAGTGTGGAACGGTGGCAATGACATCCTGG 1039

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
          |  ||||  ||||  ||  ||||||||  |  |  |  ||||||||  | | |||||||
Db   1040 CTGATCCCCAAGATGTCGCCATGCTGCTTCCCAAACTCTCCAACCTCCTGTTCCATAAGG 1099

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
          | |||| ||  |    |||| |||| ||||||||||   |||||||  |||||     | |
Db   1100 AGATTCTTGCCTACAATCACCTGGACTTCATCTGGGCAATGGATGCCCCTCAAGAGGTTT 1159

Qy   1199 acaatgaaatcatccatctgatgcagcaggaggagaccaacc 1240
          |||||||  || ||      |||||     | ||  | |  || |
Db   1160 ACAATGAGATGATTTCCATGATGGCAGAAGACTAAAAGAATC 1201
```

```
RESULT  12
RNLIP
LOCUS       RNLIP        1355 bp    mRNA            ROD       30-MAR-1995
DEFINITION  Rat mRNA for lingual lipase.
ACCESSION   X02309
VERSION     X02309.1  GI:56595
KEYWORDS    glycoprotein; lipase.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1355)
  AUTHORS   Docherty,A.J., Bodmer,M.W., Angal,S., Verger,R., Riviere,C.,
            Lowe,P.A., Lyons,A., Emtage,J.S. and Harris,T.J.
  TITLE     Molecular cloning and nucleotide sequence of rat lingual lipase
            cDNA
  JOURNAL   Nucleic Acids Res. 13 (6), 1891-1903 (1985)
  MEDLINE   85215587
COMMENT     Data kindly reviewed (12-FEB-1986) by A.J.P. Docherty.
FEATURES             Location/Qualifiers
     source          1. .1355
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     sig_peptide     49. .102
                     /product="put. signal peptide (aa -18 to -1)"
     CDS             49. .1236
                     /note="lipase precursor"
                     /codon_start=1
                     /protein_id="CAA26179.1"
                     /db_xref="GI:56596"
                     /db_xref="SWISS-PROT:P04634"
                     /translation="MWLLLITSVISTFGGAHGLFGKLGPGNPEANMNISQMITYWGYP
                     CQEYEVVTEDGYILGVYRIPHGKNNSENIGKRPVVYLQHGLIASATNWIANLPNNSLA
                     FMLADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATINFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPTLAKKIKTFYALAPVATVKYTQSPLKKISFIPT
                     FLFKLMFGKKMFLPHTYFDDFLGTEVCSREVLDLLCSNTLFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSVQDFLHWAQLVRSGKFQAFNWGSPSQNMLHYNQKTPPEYDVSAMTVPV
                     AVWNGGNDILADPQDVAMLLPKLSNLLFHKEILAYNHLDFIWAMDAPQEVYNEMISMM
                     AED"
     mat_peptide     103. .1233
                     /product="lipase (aa 1-377)"
     misc_feature    505. .507
                     /note="aa 153 pot. lipid binding site"
     misc_feature    1355
                     /note="polyA site"
BASE COUNT      400 a    291 c    276 g    387 t      1 others
ORIGIN


  Query Match           36.8%;  Score 467.6;  DB 10;  Length 1355;
  Best Local Similarity   63.5%;  Pred. No. 7.6e-129;
  Matches  713;  Conservative    0;  Mismatches  409;  Indels    0;  Gaps    0;

Qy     119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
           |||   |||| ||||||   |||||||||||| | || ||     |    || ||||| |
Db     122 CTGGAAACCCTGAAGCAAATATGAATATTAGTCAGATGATAACTTACTGGGGATATCCAT 181

Qy     179 gtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctc 238
           || | |||||||||||  |  |||||||||||| || || ||    ||  |||| |||||||
Db     182 GTCAAGAATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTACAGAATTCCTC 241

Qy     239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
             ||         |  || | || | |||    || ||||||||||   ||||||||||
Db     242 ATGGGAAGAATAATTCTGAAAATATAGGCAAGAGACCTGTGGTGTATTTGCAGCATGGTT 301

Qy     299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
           |  | | |   || |  |||||||||  | || || || ||||| ||||||| ||| || |
Db     302 TGATCGCATCAGCCACAAACTGGATTGCAAATCTACCAAACAACAGCCTGGCCTTTATGC 361

Qy     359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
           | ||||| ||||| | ||| |||||| |||||||||  | |||||  | |||||  | |
Db     362 TAGCAGACGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGAAATACATGGTCCAGGA 421

Qy     419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
           || |         |||  ||||  || ||| |||||||||||||| | |||||| ||||
Db     422 AAAATGTATACTACTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 481

Qy     479 ctaggtttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaaga 538
           |||  | |||||||||  ||    ||||||||  ||| | ||||||||  || || || ||||
Db     482 CTAAATATGACCTTCCCGCCACAATAAACTTCATTGTACAGAAAACTGGACAAGAGAAGA 541

Qy     539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
           |  | ||||| ||  |||| ||||||||||| || || || ||||| ||||| |||| |
Db     542 TACACTATGTTGGTCATTCTCAGGGCACCACTATTGGTTTCATTGCCTTTTCTACCAATC 601

Qy     599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
           |    ||||| | |||||||| | || || ||| ||||| ||  | || || || |||
Db     602 CTACACTGGCCAAAAAAATCAAGACGTTTTATGCATTAGCTCCAGTTGCTACCGTGAAGT 661

Qy     659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
           || || |||| ||| ||  ||  ||| ||  | ||      |  |  |||||   |||
Db     662 ATACACAAAGTCCCTTGAAAAAGATTTCATTTATTCCTACATTTCTTTTCAAGCTTATGT 721

Qy     719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 778
           ||||||| |||   || |||   || ||| ' |||          |||| ||   |  | |
Db     722 TTGGCAAGAAAATGTTCCTGCCCCACACCTACTTTGATGACTTTCTTGGTACCGAAGTGT 781

Qy     779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
           |    | || | |||| ||||   | || || || |   | ||  |  |  |||||||
Db     782 GCTCACGGGAGGTTCTAGATCTTCTCTGCAGCAACACTTTATTCATCTTCTGTGGATTTG 841

Qy     839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
           |||  || ||  | ||  |||| ||      ||| |||   |  || | ||  || || |
Db     842 ACAAGAAAAACTTAAATGTGAGTCGTTTTGATGTGTATCTAGGGCATAATCCAGCAGGGA 901

Qy     899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
           ||||||| ||| |  |||| ||||||   |||   || |  |||||  || | |  || |
Db     902 CATCTGTTCAAGACTTTCTCCACTGGGCACAGCTTGTTAGATCTGGGAAATTTCAAGCCT 961

Qy     959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
           |  ||||||| ||      || | ||  ||  | | | ||| |||  ||| |||      |
Db     962 TCAACTGGGGAAGCCCATCCCAGAACATGTTACACTACAACCAGAAAACGCCTCCTGAAT 1021

Qy    1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
           |    ||  ||  ||||| || ||    ||| ||||||  || ||  | |||  ||
Db    1022 ATGATGTGTCAGCCATGACTGTGCCAGTTGCAGTGTGGAACGGTGGCAATGACATCCTGG 1081

Qy    1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
           |  |||| |||| ||    ||||||||  |  |  |  |||||||| | | |||||||
Db    1082 CTGATCCCCAAGATGTCGCCATGCTGCTTCCCAAACTCTCCAACCTCCTGTTCCATAAGG 1141

Qy    1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
           | |||| ||  |    |||| |||| ||||||||||   ||||||| |||||     | |
Db    1142 AGATTCTTGCCTACAATCACCTGGACTTCATCTGGGCAATGGATGCCCCTCAAGAGGTTT 1201

Qy    1199 acaatgaaatcatccatctgatgcagcaggaggagaccaacc 1240
           ||||||| || ||     |||||    | ||  | |  || |
Db    1202 ACAATGAGATGATTTCCATGATGGCAGAAGACTAAAAGAATC 1243


RESULT  13
MMLYACLY
LOCUS       MMLYACLY     2358 bp    mRNA            ROD       24-FEB-1999
DEFINITION  M.musculus (C57 Black/6X CBA) LAL mRNA for lysosomal acid lipase.
ACCESSION   Z31689
VERSION     Z31689.1  GI:4456670
KEYWORDS    LAL; lipase; lysosomal acid lipase.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2358)
  AUTHORS   Du,H., Witte,D.P. and Grabowski,G.A.
```

```
  TITLE     Tissue and cellular specific expression of murine lysosomal acid
            lipase mRNA and protein
  JOURNAL   J. Lipid Res. 37 (5), 937-949 (1996)
  MEDLINE   96363957
REFERENCE   2  (bases 1 to 2358)
  AUTHORS   Du,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1994) Hong Du, Division of Human Genetics,
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH, 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2358
                     /organism="Mus musculus"
                     /strain="C57 Black/6X CBA"
                     /db_xref="taxon:10090"
                     /sex="Female"
                     /clone="mLAL #49"
                     /tissue_type="liver"
                     /clone_lib="Mouse liver cDNA library /Lamda-ZAP"
     gene            19. .1212
                     /gene="LAL"
     CDS             19. .1212
                     /gene="LAL"
                     /citation=[1]
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="CAA83494.1"
                     /db_xref="GI:4456671"
                     /translation="MQLQGLVFVFTIGILLSRVPTGTVSAVDPEVNMNVTEIIMRWGY
                     PGEEHSVLTGDGYILSIHRIPRGWKNHFGKGPRPVVYLQHGLLADSSNWVTNIDNSSL
                     GFLLADRGFDVWMGNSRGNTWSLKHKTLSVSQDEFWAFSFDEMAKYDLPASINYILNK
                     TGQEQIYYVGHSQGCTIGFIAFSQMPELAKKIKMFLVLAPVLSLNFASGPLLQLGRLP
                     DPLLKDMFGQKQFLPQSAMLKWLSIHVCTHVIMKELCANVFFLLCGFNEKNLNMSRVD
                     VYTTHCPAELLVQNMLHWGQVFKYRKLQAFDWGSSEKNYFHYNQSFPPSYNIKNMRLP
                     TALWSGGRDWLADINDITILLTQIPKLVYHKNIPEWDHLDFIWGLDAPWKLYDEIISL
                     MKKYQ"
     sig_peptide     19. .75
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     mat_peptide     76. .1209
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     polyA_signal    2246. .2251
     polyA_site      2358
BASE COUNT      604 a    549 c    524 g    681 t
ORIGIN

  Query Match           36.3%;  Score 461;  DB 10;  Length 2358;
  Best Local Similarity  63.3%;  Pred. No. 7.8e-127;
  Matches  707;  Conservative    0;  Mismatches  410;  Indels    0;  Gaps    0;

Qy    117 agctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatcc 176
          ||||||||| || |||| |||||| | |  || || |||  |    || |||||
Db     93 AGCTGTGGATCCTGAAGTCAACATGAATGTGACCGAGATAATCATGCGCTGGGGATATCC 152

Qy    177 ctgtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcc 236
          |||||||| |    |||  || | |||||| ||||| ||  |  || || | || ||
Db    153 TGGTGAGGAACACTCGGTCCTGACAGGAGATGGCTATATTCTGAGTATTCACCGAATCCC 212

Qy    237 tcgaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatgg 296
          ||| ||      | | | |      | |||| |||| || || ||||  || ||||| ||
Db    213 TCGTGGGTGGAAGAACCATTTTGGCAAAGGTCCCAGACCAGTTGTGTATCTTCAGCACGG 272

Qy    297 cctagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcat 356
          | |  | | || | |||| |||||| |  | ||| |   |||||  ||| |||||||| |
Db    273 CTTGCTGGCAGATTCTAGTAACTGGGTCACAAACATTGACAACAGCAGCTTGGGCTTCCT 332

Qy    357 tctggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcg 416
          |||||  |||  ||| ||||| |||||||||| |||||||| |||||| |||||||||
Db    333 CCTGGCGGATCGTGGCTTTGATGTGTGGATGGGAAACAGCAGAGGAAACACCTGGTCTCT 392

Qy    417 aaaacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagat 476
          || |||||||| || ||  |    || |||||||||||||| ||||||| |||||||||
Db    393 GAAGCACAAGACTCTTTCAGTTTCTCAGGATGAGTTCTGGGCCTTCAGTTTTGATGAGAT 452

Qy    477 ggctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaa 536
          ||| |  | |||||| |||||    || ||||  |||||| | ||||| ||||| ||| |
Db    453 GGCAAAATATGACCTACCTGCTTCAATTAACTACATTTTGAATAAAACTGGCCAAGAACA 512

Qy    537 gatctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccat 596
          ||||||||||| ||  |||| || |||  |||||| || || || || |||||    ||
Db    513 AATCTATTATGTGGGTCATTCTCAAGGCTGCACCATAGGTTTCATAGCTTTTTCACAGAT 572

Qy    597 gccagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaa 656
          ||| ||| |||| |||| || ||||||| ||| |  || || ||| |     || | ||
Db    573 GCCTGAGTTGGCAAAGAAGATTAAAATGTTTTTAGTCTTGGCTCCCGTGTTGTCTCTCAA 632

Qy    657 gcatgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggatt 716
            |||   | ||||       ||||       |||||||     |  ||||||   |
Db    633 TTTTGCTTCAGGCCCGCTACTTCAATTGGGACGCTTGCCAGACCCTCTTCTCAAGGACAT 692

Qy    717 gtttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacct 776
          ||||||  | ||  |||||||   |||||      | || |     ||    ||| || |
Db    693 GTTTGGACAGAAGCAATTTCTCCCTCAGAGTGCGATGCTGAAGTGGCTGAGCATTCACGT 752

Qy    777 ttgtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggatt 836
          |||   ||| || ||  |  |  || ||||    ||  || | || || ||  |||| ||
Db    753 TTGCACCCACGTCATCATGAAGGAGCTTTGCGCAAACGTCTTCTTCCTGCTATGTGGCTT 812

Qy    837 caacaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctgg 896
          |||    || ||| | || |||    |||    ||| ||| |  | |||   | ||| |
Db    813 CAATGAGAAGAATTTAAATATGTCTAGAGTGGATGTGTATACAACACACTGCCCTGCGGA 872

Qy    897 aacatctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggc 956
          |      |||||||| ||  |||||||| | ||||   | || |   || | || |  ||
Db    873 ACTTCTCGTGCAAAACATGTTACACTGGGGTCAGGTTTTTAAATACCGTAAGCTTCAAGC 932

Qy    957 atttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaag 1016
          |||||||||||  ||       || |||       | | ||| |||     ||||
Db    933 TTTTGACTGGGGAAGCAGTGAAAAGAATTATTTTCATTACAACCAGAGTTTTCCTCCCTC 992

Qy   1017 gtacagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggct 1076
          |||||   | | | | |||    | || || ||  ||||||  || || | |||||||||
Db    993 GTACAACATAAAAAACATGCGTCTACCCACTGCTTTGTGGAGCGGGGGCCGGGACTGGCT 1052

Qy   1077 ttcaaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataa 1136
            || |     | ||| | |  ||  | ||  || | |  |||| |  | || || ||
Db   1053 GGCAGACATCAATGACATCACTATCTTACTGACTCAAATCCCCAAGTTGGTGTATCACAA 1112

Qy   1137 gaatattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtat 1196
          ||| |||||||| |||| ||  ||||||| ||||||||||||||||| ||        |
Db   1113 GAACATTCCTGAGTGGGACCATCTGGATTTTATCTGGGGTTTGGATGCCCCCTGGAAACT 1172

Qy   1197 gtacaatgaaatcatccatctgatgcagcaggaggag 1233
          |||  |||||| |||  ||| ||| || |  |  ||
Db   1173 GTATGATGAAATAATCAGTCTAATGAAGAAATACCAG 1209


RESULT  14
S81497
LOCUS       S81497       3144 bp    mRNA            ROD       02-AUG-2000
DEFINITION  lysosomal acid lipase-intracellular hydrolase [rats, Wolman, liver,
            mRNA, 3144 nt].
ACCESSION   S81497
VERSION     S81497.1  GI:1336725
KEYWORDS    .
SOURCE      Rattus sp. liver Wolman.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 3144)
  AUTHORS   Nakagawa,H., Matsubara,S., Kuriyama,M., Yoshidome,H., Fujiyama,J.,
            Yoshida,H. and Osame,M.
  TITLE     Cloning of rat lysosomal acid lipase cDNA and identification of the
```

```
          mutation in the rat model of Wolman's disease
  JOURNAL   J. Lipid Res. 36 (10), 2212-2218 (1995)
  MEDLINE   96129534
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 176569] from the original journal article.
            This sequence comes from Fig. 2.
FEATURES             Location/Qualifiers
     source          1. .3144
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
     gene            1. .3144
                     /gene="lysosomal acid lipase, LAL"
     CDS             233. .1426
                     /gene="lysosomal acid lipase, LAL"
                     /note="intracellular hydrolase; This sequence comes from
                     Fig. 2; conceptual translation presented here differs from
                     translation in publication; LAL"
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="AAB36043.2"
                     /db_xref="GI:9653291"
                     /translation="MQLLGRVICFVVGILLSGGPTGTISAVDPEANMNVTEIIMHWGY
                     PEHSVQTGDGYILGVHRIPHGRKNQFDKGPKPVVYLQWRHGFLADSSNWVTNIDNNSL
                     GFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEYWAFSFDEMAKYDLPASINYILNK
                     TGQEQLYNVGHSQGCTIGFIAFSQMPELAKKVKMFFALAPVLSLNFASGPMVKLGRLP
                     DLLLEDLFGQKQFLPQSAMVKWLSTHICTHVIMKELCANIFFLICGFNEKNLNMSRVD
                     VYTTHCPAGTSVQNMVHWTQVVKYHKLQAFDWGSSDKNYFHYNQSYPPLYSIKDMQLP
                     TALWSGGKDWLADTSDINILLTEIPTLVYHKNIPEWDHLDFIWGLDAPWRLYNEVVSL
                     MKKYQ"
BASE COUNT      852 a    690 c    754 g    848 t
ORIGIN
```

Query Match 36.2%; Score 459.8; DB 10; Length 3144;
Best Local Similarity 64.5%; Pred. No. 1.9e-126;
Matches 724; Conservative 0; Mismatches 387; Indels 12; Gaps 2;

```
Qy    117 agctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatcc 176
Db    307 AGCTGTGGACCCTGAAGCAAACATGAATGTGACTGAGATAATCATGCACTGGGGATATCC 366

Qy    177 ctgtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcc 236
Db    367 A------GAGCACTCAGTACAGACAGGAGATGGCTACATTCTGGGTGTTCACCGAATCCC 420

Qy    237 tcgaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcag----- 291
Db    421 GCATGGGCGGAAGAACCAGTTTGATAAAGGTCCCAAACCAGTTGTGTATCTTCAGTGGCG 480

Qy    292 -catggcctagttggaggtgctagcaactggatttccaacctgcccaacaatagcctggg 350
Db    481 ACATGGCTTCCTGGCAGATTCTAGTAACTGGGTCACAAACATTGACAACAACAGCCTGGG 540

Qy    351 cttcattctggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctg 410
Db    541 CTTCATCCTGGCAGATGCTGGCTTTGATGTGTGGATGGGAAACAGCAGAGGAAACACCTG 600

Qy    411 gtctcgaaaacacaagacactctccatagaccaagatgagttctgggctttcagttatga 470
Db    601 GTCTCGGAAGCACAAGACTCTGTCAGTTTCTCAGGATGAATACTGGGCCTTCAGTTTTGA 660

Qy    471 tgagatggctaggtttgaccttcctgcagtgataaactttatttttgcagaaaacgggcca 530
Db    661 TGAGATGGCAAAATATGACCTCCCTGCTTCAATTAACTACATTTTGAATAAAACTGGCCA 720

Qy    531 ggaaaagatctattatgtcggctattcacagggcaccaccatgggctttattgcattttc 590
Db    721 AGAACAGCTGTATAATGTGGGCCACTCTCAAGGCTGCACCATAGGTTTTATAGCATTTTC 780

Qy    591 caccatgccagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccac 650
Db    781 ACAGATGCCTGAACTGGCAAAGAAGGTTAAAATGTTTTTTGCCCTGGCTCCTGTGTTGTC 840
```

```
Qy    651 tgttaagcatgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaa 710
Db    841 TCTCAATTTTGCTTCGGGCCCGATGGTCAAATTGGGACGGTTGCCGGATCTGCTTCTTGA 900

Qy    711 gggattgtttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttat 770
Db    901 GGACTTGTTTGGGCAGAAGCAATTTCTTCCTCAGAGTGCAATGGTGAAATGGCTAAGCAC 960

Qy    771 ttacctttgtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctggg 830
Db    961 TCACATTTGCACGCACGTCATCATGAAGGAGCTTTGCGCCAACATCTTCTTTCTGATATG 1020

Qy    831 tggattcaacaccaacaatatgaacatgagtcgagcaagtgtatatgctgcccacactct 890
Db   1021 TGGCTTCAATGAGAAGAATTTAAACATGTCTAGAGTGGATGTGTATACAACACACTGTCC 1080

Qy    891 tgctggaacatctgtgcaaaatattctacactggagccaggcagtgaattctggtgaact 950
Db   1081 TGCGGGAACCTCTGTGCAAAACATGGTACACTGGACCCAGGTTGTTAAATACCATAAGCT 1140

Qy    951 ccgggcatttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcc 1010
Db   1141 TCAAGCTTTTGACTGGGGAAGCAGTGACAAGAATTATTTTCATTACAATCAGAGTTATCC 1200

Qy   1011 tgtaaggtacagagtcagagatatgacggtccctacagcaatgtggacaggaggtcagga 1070
Db   1201 ACCCTTGTACAGCATAAAAGACATGCAACTACCCACTGCTTTATGGAGCGGGGGCAAGGA 1260

Qy   1071 ctggctttcaaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatcta 1130
Db   1261 CTGGCTGGCAGACACCAGTGACATCAATATCTTACTGACTGAAATCCCCACGTTAGTGTA 1320

Qy   1131 ccataagaatattcctgaatgggctcacgtggatttcatctggggtttggatgctcctca 1190
Db   1321 CCACAAGAACATTCCCGAGTGGGACCATCTGGATTTTATCTGGGGTTTGGATGCCCCCTG 1380

Qy   1191 ccgtatgtacaatgaaatcatccatctgatgcagcaggaggag 1233
Db   1381 GAGGCTGTATAATGAAGTAGTCAGTCTAATGAAGAAATACCAG 1423
```

```
RESULT  15
BOVMPE
LOCUS       BOVMPE      1417 bp    mRNA            MAM       10-DEC-1994
DEFINITION  Bos taurus pregastric esterase mRNA, complete cds.
ACCESSION   L26319
VERSION     L26319.1  GI:600756
KEYWORDS    esterase; pregastric esterase.
SOURCE      Bos taurus 25 weeks Tongue serous gland cDNA to mRNA.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1417)
  AUTHORS   Timmermans,M.Y.J., Kupers,L.P. and Teuchy,H.
  TITLE     The cDNA sequence encoding bovine pregastric esterase
  JOURNAL   Gene 147, 259-262 (1994)
  MEDLINE   95011625
FEATURES             Location/Qualifiers
     source          1. .1417
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /cell_type="Secretory cell"
                     /dev_stage="25 weeks"
                     /tissue_type="Tongue serous gland"
     sig_peptide     73. .129
     CDS             73. .1266
                     /EC_number="3.1.1.3"
                     /codon_start=1
                     /product="pregastric esterase"
                     /protein_id="AAA57037.1"
                     /db_xref="GI:600757"
                     /translation="MWWLLVTVCFIHMSGNAFCFLGKIAKNPEASMNVSQMISYWGYP
```

```
                     SEMHKVITADGYILQVYRIPHGKNNANHLGQRPVVFLQHGLLGSATNWISNLPKNSLG
                     FLLADAGYDVWLGNSRGNTWAQEHLYYSPDSPEFWAFSFDEMAEYDLPSTIDFILRRT
                     GQKKLHYVGHSQGTTIGFIAFSTSPTLAEKIKVFYALAPVATVKYTKSLFNKLALIPH
                     FLFKIIFGDKMFYPHTFLEQFLGVEMCSRETLDVLCKNALFAITGVDNKNFNMSRLDV
                     YIAHNPAGTSVQNTLHWRQAVKSGKFQAFDWGAPYQNLMHYHQPTPPIYNLTAMNVPI
                     AVWSADNDLLADPQDVDFLLSKLSNLIYHKEIPNYNHLDFIWAMDAPQEVYNEIVSLM
                     AEDKK"
     mat_peptide     130. .1263
                     /EC_number="3.1.1.3"
                     /function="triacylglycerol acylhydrolase"
                     /evidence=experimental
                     /product="pregastric esterase"
     polyA_signal    1378. .1385
     polyA_site      1396
BASE COUNT      408 a    306 c    276 g    427 t
ORIGIN
```

```
  Query Match           36.1%;  Score 458;  DB 4;  Length 1417;
  Best Local Similarity  62.1%;  Pred. No. 5.8e-126;
  Matches  741;  Conservative    0;  Mismatches  450;  Indels    3;  Gaps    1;

Qy      44 tggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattcagtac 103
           || ||||||||   ||  || |    | || | |  |||| |    ||  ||| || |
Db      68 TGAAAATGTGGTGGCTACTTGTAACAGTGTGTTTCATCCACATGTCTGGAAATGCATTTT 127

Qy     104 atatgc---caactaaagctgtggacccagaagcattcatgaatattagtgaaatcatcc 160
           | | |    ||  |  |||  | |||| |||||    |||||| ||||| | || ||
Db     128 GTTTCCTTGGAAAAATTGCTAAGAACCCTGAAGCCAGTATGAATGTTAGTCAGATGATTT 187

Qy     161 aacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggtatatccttt 220
           |    ||||| ||  ||||||    || ||||   |||||  ||||| |||||||||
Db     188 CCTACTGGGGCTACCCAAGTGAGATGCATAAAGTTATAACTGCGGATGGTTATATCCTTC 247

Qy     221 ctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccaggcctgtgg 280
           ||  | ||||||||||  ||        |  ||||  |   ||||   || ||||| |
Db     248 AGGTCTATCGGATTCCTCATGGAAAGAATAATGCTAATCATTTAGGTCAGAGACCTGTTG 307

Qy     281 tgttactgcagcatggcctagttggaggtgctagcaactggatttccaacctgcccaaca 340
           |||| ||||||||||| ||  ||||||   || |  |||||||||||||||||||||| |
Db     308 TGTTTCTGCAGCATGGTCTTCTTGGATCAGCCACAAACTGGATTTCCAACCTGCCCAAGA 367

Qy     341 atagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaacagcaggg 400
           | |||||||||||| | ||||||||||||||||| |||||||||| ||||||||||||| |
Db     368 ACAGCCTGGGCTTCCTCCTGGCAGATGCTGGTTATGACGTGTGGCTGGGGAACAGCAGAG 427

Qy     401 gaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagttctgggctt 460
           |||||  |||||  | |   ||||          ||   ||||      || ||||||||||
Db     428 GAAACACCTGGGCCCAGGAACATTTATACTATTCACCAGACTCCCCGGAATTCTGGGCTT 487

Qy     461 tcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttatttttgcaga 520
           |||| | ||||||  |||||    | |||||||||  |    ||  | || || ||   ||
Db     488 TCAGCTTTGATGAAATGGCGGAATATGACCTTCCATCTACAATTGATTTCATCTTAAGGA 547

Qy     521 aaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccatgggcttta 580
           ||| || ||| | ||| |  | ||||| ||| |||| || ||||||||||| || ||||
Db     548 GAACAGGACAGAAGAAGCTACACTATGTTGGCCATTCCCAAGGCACCACCATTGGTTTTA 607

Qy     581 ttgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgctttagcac 640
           | || ||||| ||||  ||     ||||| | |||||||||| | |  | ||| ||||| |
Db     608 TCGCCTTTTCTACCAGTCCCACATTGGCTGAAAAAATCAAAGTCTTCTATGCATTAGCCC 667

Qy     641 ccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgctgccagata 700
           |  | ||||| || ||| |  | || ||||     | |||| ||    |  | ||  |
Db     668 CAGTTGCCACAGTGAAGTACACCAAGAGCCTGTTTAACAAACTTGCACTTATTCCTCACT 727

Qy     701 tgatgatcaaggattgtttggcaaaaaagaatttctgtatcagaccagatttctcagac 760
           |  |  |||||     | |||||  | |||   ||        || ||    ||      |
Db     728 TCCTCTTCAAGATTATATTTGGTGACAAAATGTTCTACCCACACACTTTTTTGGAACAAT 787

Qy     761 aacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaatatcatgt 820
           ||||  | || |  | ||   ||  | ||  || |||    ||||||  |||  | |||
Db     788 TTCTTGGTGTTGAAATGTGCTCCCGTGAGACACTGGATGTCCTTTGTAAGAATGCCTTGT 847
```

```
Qy     821 tacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgtatatgctg 880
           |     |  |||| |  |||  || ||  | |||||||| ||    |  ||| |||  |
Db     848 TTGCCATTACTGGAGTTGACAATAAAAACTTCAACATGAGTCGCTTAGATGTGTATATAG 907

Qy     881 cccacactcttgctggaacatctgtgcaaaatattctacactggagccaggcagtgaatt 940
           | || | ||  || ||||| ||||| ||||| |  || ||||||||| ||||| || || |
Db     908 CACATAATCCAGCAGGAACTTCTGTTCAAAACACCCTCCACTGGAGACAGGCTGTTAAGT 967

Qy     941 ctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaaatgcaatc 1000
           ||||  || |||  || |||||||||||          | || ||     | |   |||
Db     968 CTGGGAAATTCCAAGCTTTTGACTGGGGAGCCCCATATCAGAACCTAATGCATTATCATC 1027

Qy    1001 agccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaatgtggacag 1060
           |||| || |||   |  ||||   | | ||  ||||  ||||| |  ||| | ||||  |
Db    1028 AGCCCACACCTCCCATCTACAATTTAACAGCCATGAATGTCCCAATTGCAGTATGGAGTG 1087

Qy    1061 gaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctgaggtgacca 1120
           |   | |||  | |  |  | ||  | || ||  |  | ||||| ||  |  |  | |
Db    1088 CTGACAATGACCTGTTGGCTGACCCTCAGGATGTTGACTTTCTGCTTTCAAAACTCTCTA 1147

Qy    1121 acctcatctaccataagaatattcctgaatgggctcacgtggatttcatctggggtttgg 1180
           | |||| |||| ||| | ||||  | |    |||| |||| || |||||||   |||
Db    1148 ATCTCATTTACCACAAGGAAATTCCAAATTACAATCACTTGGACTTTATCTGGGCAATGG 1207

Qy    1181 atgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggaggaga 1234
           |||| |||||     | ||||||||||  |   | |||||    | ||  | |
Db    1208 ATGCACCTCAAGAAGTTTACAATGAAATTGTTTCTTTGATGGCCGAAGACAAAA 1261
```

Search completed: January 30, 2002, 22:09:16
Job time: 6263 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 30, 2002, 21:08:53 ; Search time 166.02 Seconds
(without alignments)
6553.100 Million cell updates/sec

Title: US-09-578-063-46
Perfect score: 1269
Sequence: 1 atgttggaaaccttgtcaag.....gacggtgtgaggccgtattg 1269

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1269	100.0	1269	22	AAF45132	Human TANGO 294 OR
2	1269	100.0	2044	22	AAF45131	Human TANGO 294 OR
3	526.2	41.5	2493	22	AAH02906	Human shear stress
4	489.2	38.6	1378	14	AAQ42310	RGL precursor. Or
5	470	37.0	1367	7	AAN60685	Sequence encoding
6	470	37.0	1367	7	AAN60566	Sequence encoding
7	470	37.0	1367	17	AAT58916	Human gastric lipa
8	468.4	36.9	1695	22	AAH57420	Human stomach cell
9	464.4	36.6	1336	6	AAN50385	Rat prelingual lip
10	444.2	35.0	1528	15	AAQ68388	Canine gastric lip
11	444.2	35.0	1528	17	AAT58915	Dog gastric lipase
12	444.2	35.0	1531	15	AAQ68389	Canine gastric lip
13	286	22.5	699	22	AAF28679	Human protein HP03
14	286	22.5	1308	22	AAF28689	Human protein HP03
15	257.8	20.3	617	21	AAA44349	Human secreted exp
16	193.2	15.2	572	21	AAA43744	Mouse secreted exp
c 17	100.2	7.9	402	22	AAH57212	Human stomach spec
18	100	7.9	1718	21	AAZ34958	Soybean acid triac
19	89.4	7.0	1483	21	AAZ34956	Rice acid triacylg
20	84.8	6.7	859	21	AAZ34953	Corn acid triacylg
21	72	5.7	1438	21	AAZ34959	Soybean acid triac
c 22	68	5.4	936	22	AAF58252	Oligonucleotide D1
c 23	68	5.4	936	22	AAF58254	Oligonucleotide D1
c 24	68	5.4	936	22	AAF58257	Oligonucleotide D1
c 25	68	5.4	936	22	AAF58259	Oligonucleotide D2
c 26	68	5.4	936	22	AAF58262	Oligonucleotide D2
c 27	68	5.4	938	22	AAF58255	Oligonucleotide D1
28	66.2	5.2	936	22	AAF58252	Oligonucleotide D1
29	66.2	5.2	936	22	AAF58254	Oligonucleotide D1
30	66.2	5.2	936	22	AAF58257	Oligonucleotide D1
31	66.2	5.2	936	22	AAF58259	Oligonucleotide D2
32	66.2	5.2	936	22	AAF58262	Oligonucleotide D2
33	66.2	5.2	938	22	AAF58255	Oligonucleotide D1
34	55.8	4.4	705	21	AAZ34952	Catalpa acid triac
35	54	4.3	549	21	AAC93827	Cat flea hindgut a
36	51	4.0	509	21	AAZ34954	Corn acid triacylg
37	48.2	3.8	178	19	AAV35431	Cell death regulat
38	46.4	3.7	43676	21	AAZ60606	Nucleotide sequenc
c 39	38.6	3.0	1664976	19	AAV21209	Methanococcus jann
c 40	36.8	2.9	13254	21	AAA32165	Complete nucleotid
c 41	36.6	2.9	244	22	AAF58238	Oligonucleotide D1
42	36.6	2.9	3271	19	AAV00060	Human G-protein co
43	36.6	2.9	3282	17	AAT40914	DNA encoding VH re
44	36.6	2.9	3282	21	AAA31024	Modified heavy cha
c 45	36.6	2.9	3282	21	AAA31038	B12 IgG1 preferred

ALIGNMENTS

RESULT 1
AAF45132
ID AAF45132 standard; cDNA; 1269 BP.
XX
AC AAF45132;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294 ORF.
XX
KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety; ss.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
DR P-PSDB; AAB66065.
XX

PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease -
XX
PS Claim 1; Fig 6; 359pp; English.
XX
CC The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other
CC neurological and cerebrovascular disorders. The CNS disorders include
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC autonomic function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder.
XX
SQ Sequence 1269 BP; 358 A; 268 C; 309 G; 334 T; 0 other;

Query Match 100.0%; Score 1269; DB 22; Length 1269;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 60

Qy     61 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 120

Qy    121 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 180

Qy    181 gaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcga 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 gaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcga 240

Qy    241 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 300

Qy    301 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 360

Qy    361 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 420

Qy    421 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 480

Qy    481 aggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 aggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatc 540

Qy    541 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 600

Qy    601 gagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 gagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcat 660

Qy    661 gcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgttt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgttt 720

Qy    721 ggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgt 780

Qy    781 ggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaac 840

Qy    841 accaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaaca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 accaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaaca 900

Qy    901 tctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcattt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 tctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcattt 960

Qy    961 gactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtac 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 gactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtac 1020

Qy   1021 agagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttca 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 agagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttca 1080

Qy   1081 aatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 aatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaat 1140

Qy   1141 attcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtac 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 attcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtac 1200

Qy   1201 aatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggtgtgag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 aatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggtgtgag 1260

Qy   1261 gccgtattg 1269
          |||||||||
Db   1261 gccgtattg 1269
```

RESULT 2
AAF45131
ID AAF45131 standard; cDNA; 2044 BP.
XX
AC AAF45131;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294 cDNA.
XX
KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety; ss.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;

```
XX
DR   WPI; 2001-032313/04.
DR   P-PSDB; AAB66065.
XX
PT   TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT   screening assays and diagnostic assays and for the treatment of
PT   neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT   disease -
XX
PS   Claim 1; Fig 6; 359pp; English.
XX
CC   The present invention relates to TANGO or INTERCEPT proteins and coding
CC   sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC   AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC   sequences are useful for the treatment of neurological disorders such as
CC   central nervous system (CNS) disorders, CNS-related disorders, focal
CC   brain disorders, global-diffuse cerebral disorders and other
CC   neurological and cerebrovascular disorders. The CNS disorders include
CC   Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC   lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC   autonomic function disorders such as hypertension and sleep disorders,
CC   neuropsychiatric disorders, psychoactive substance use disorders,
CC   anxiety, and bipolar affective disorder.
XX
SQ   Sequence 2044 BP; 618 A; 401 C; 460 G; 565 T; 0 other;

  Query Match             100.0%;  Score 1269;  DB 22;  Length 2044;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1269;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 185

Qy      61 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     186 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 245

Qy     121 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     246 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 305

Qy     181 gaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcga 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     306 gaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcga 365

Qy     241 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     366 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 425

Qy     301 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     426 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 485

Qy     361 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     486 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 545

Qy     421 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     546 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 605

Qy     481 aggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     606 aggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatc 665

Qy     541 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     666 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 725

Qy     601 gagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcat 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     726 gagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcat 785

Qy     661 gcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgttt 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     786 gcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgttt 845

Qy     721 ggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgt 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     846 ggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgt 905

Qy     781 ggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaac 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     906 ggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaac 965

Qy     841 accaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaaca 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     966 accaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaaca 1025

Qy     901 tctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcattt 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1026 tctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcattt 1085

Qy     961 gactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtac 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1086 gactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtac 1145

Qy    1021 agagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttca 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1146 agagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttca 1205

Qy    1081 aatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaat 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1206 aatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaat 1265

Qy    1141 attcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtac 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1266 attcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtac 1325

Qy    1201 aatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggtgtgag 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1326 aatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggtgtgag 1385

Qy    1261 gccgtattg 1269
           |||||||||
Db    1386 gccgtattg 1394


RESULT   3
AAH02906
ID   AAH02906 standard; DNA; 2493 BP.
XX
AC   AAH02906;
XX
DT   15-JUN-2001  (first entry)
XX
DE   Human shear stress-response coding sequence SEQ ID NO: 65.
XX
KW   Human; shear stress-response protein; vascular disease;
KW   arteriosclerosis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200125427-A1.
XX
PD   12-APR-2001.
XX
PF   02-OCT-2000; 2000WO-JP06840.
XX
PR   01-OCT-1999;   99JP-0280976.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (NOJI/) NOJIMA H.
XX
```

PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K·
PI Kuga T, Sekine S, Nakamura Y, Sugano S;
XX
DR WPI; 2001-266308/27
DR P-PSDB; AAB90783.
XX
PT DNA sequences, proteins encoded by them and antibodies against them.
PT useful in diagnosis and treatment of vascular disease caused by
PT arteriosclerosis -
XX
PS Claim 20; Page 398-402; 678pp; Japanese.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human shear stress response proteins. These are useful in the
CC diagnosis, treatment and screening of vascular diseases caused by
CC arteriosclerosis, including heart failure, post-PTCA restenosis and
CC hypertension.
XX
SQ Sequence 2493 BP; 692 A; 484 C; 516 G; 801 T; 0 other;

Query Match 41.5%; Score 526.2; DB 22; Length 2493;
Best Local Similarity 65.3%; Pred. No. 8.7e-157;
Matches 790; Conservative 0; Mismatches 413; Indels 6; Gaps 1;

```
Qy      39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
           ||||||| ||||| || | ||    |||||| | | || | |   |    || ||||
Db      37 cagaatgaaaatgcggttcttggggttggtggtctgtttggttctctggccccctgcattc 96

Qy      99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152
           |       |  |   | |||||||| || ||| ||  ||||||| | |||||
Db      97 tgaggggtctggagggaaactgacagctgtggatcctgaaacaaacatgaatgtgagtga 156

Qy     153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212
           ||| |||    |    || |  ||  |||||||||   ||| |  || |||||||| ||
Db     157 aattatctcttactggggattccctagtgaggaatacctagttgagacagaagatggata 216

Qy     213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272
           ||| || |  ||||| | |||||||  ||      | | | |    | | |||| |||
Db     217 tattctgtgccttaaccgaattcctcatgggaggaagaaccattctgacaaaggtcccaa 276

Qy     273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332
           || || || || ||||| |||||| |  | | || | |||| |||||| |  | |||||
Db     277 accagttgtcttcctgcaacatggcttgctggcagattctagtaactgggtcacaaacct 336

Qy     333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392
           ||||||  |||||||||||||||||| || ||||||||||||||||||||||||||| ||
Db     337 tgccaacagcagcctgggcttcattcttgctgatgctggttttgacgtgtggatgggcaa 396

Qy     393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452
           |||||| |||||  |||||||||| ||||| ||||||||||  |    || |||||
Db     397 cagcagaggaaatacctggtctcggaaacataagacactctcagtttctcaggatgaatt 456

Qy     453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512
           |||||||||||||||||||||||||| |  | ||||| || ||    || ||||| ||
Db     457 ctgggctttcagttatgatgagatggcaaaatatgacctaccagcttccattaacttcat 516

Qy     513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572
           | || | ||||| ||||| ||| |  | |||||||| ||  |||| || |||||||| ||
Db     517 tctgaataaaactggccaagaacaagtgtattatgtgggtcattctcaaggcaccactat 576

Qy     573 gggctttattgcatttttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632
           || ||||| ||||||||    || || |||||||| | |  || ||||||| ||||||
Db     577 aggttttatagcatttttcacagatccctgagctggctaaaaggattaaaatgtttttgc 636

Qy     633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 692
           |  |  ||  | ||  | ||          |. |||||   | |||||||        |
Db     637 cctgggtcctgtggcttccgtcgccttctgtactagccctatggccaaattaggacgatt 696

Qy     693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752
           |||||||    | || ||||  || |||||  | ||||||||||    ||||     ||
Db     697 accagatcatctcattaaggacttatttggagacaaagaatttcttccccagagtgcgtt 756
```

```
Qy     753 tctcagacaacttgttatttacctttqtggccaggtgattcttg.tcaa.ttgg..tad 812
           | | |    || | ||   || ||||    || || || ||  ;  ||' | ||| | ||
Db     757 tttgaaqtggctgggtacccacgttgcactcatgtcatact.aaggaagctgtgtggaaa 816

Qy     813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccggagcaagtgt 872
           | ||   || |||||| ||||||| ||    |  ||| | || |||    |||   |||
Db     817 tctctgttttcttctgtgtggatttaatgagagaaatttaaatatgctagagtggatgt 876

Qy     873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932
           |||| |  | ||  ||| |||||||| |||||||||| ||  ||||||||||||||||
Db     877 atatacaacacattctcctgctggaacttctgtgcaaaacatgttacactggagccaggc 936

Qy     933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992
           || || |      |  | |  || |||||||||| ||     |||| |||       |
Db     937 tgttaaattccaaaagtttcaagcctttgactggggaagcagtgccaagaattatttca 996

Qy     993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052
           | ||| |||     ||||   | ||||  || |  || |||   || || || ||| |
Db     997 ttacaaccagagttatcctcccacatacaatgtgaaggacatgcttgtgccgactgcagt 1056

Qy    1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112
           ||||  || ||||| ||||||||| || ||    | ||||| || ||  | ||  || |
Db    1057 ctggagcgggggtcacgactggcttgcagatgtctacgacgtcaatatcttactgactca 1116

Qy    1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172
           | | |||||| |  | | |||| ||| ||||| ||||||| ||  | || ||||| ||
Db    1117 gatcaccaacttggtgttccatgagagcattccggaatgggagcatcttgacttcatttg 1176

Qy    1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232
           ||| ||||||| |||    |  | || ||| |||| ||  |||| ||| |  |  | |
Db    1177 gggcctggatgccccttggaggctttataataaaattattaatctaatgaggaaatatca 1236

Qy    1233 gaccaacct 1241
           |   || ||
Db    1237 gtgaaagct 1245

RESULT 4
AAQ42310
ID AAQ42310 standard; DNA; 1378 BP.
XX
AC AAQ42310;
XX
DT 20-SEP-1993 (first entry)
XX
DE RGL precursor.
XX
KW Rabbit gastric lipase; RGL; pRGLN2.1; fat; bioconversion;
KW hydrolysis; transesterification; ds.
XX
OS Oryctolagus cuniculus.
XX
FH Key Location/Qualifiers
FT misc_RNA 120..1247
FT /*tag= a
FT /note= "claim 1; page 10-11"
XX
PN EP542629-A.
XX
PD 19-MAY-1993.
XX
PF 12-NOV-1992; 92EP-0403055.
XX
PR 13-NOV-1991; 91FR-0013948.
XX
PA (LJOU ) INST RECH JOUVEINAL.
XX
PI Benicourt C, Blanchard C, Junien J;
XX
DR WPI; 1993-161080/20.
DR P-PSDB; AAR37302.
XX
PT Rabbit gastric lipase, its precursor and their DNA - useful for
```

```
PT   treating conditions linked to gastric lipase deficiency, such as
PT   mucoviscidiosis and pancreatic exocrine insufficiency
XX
PS   Claim 1; Fig 6; 31pp; French.
XX
CC   mRNA was isolated from rabbit gastric mucosal cells, converted to
CC   cDNA and fragments cloned in pUC18.  Recombinant plasmids were used
CC   to transform E.coli MM294 and cells screened for reactivity with
CC   probes corresp. to known parts of the RGL sequence.  One positive
CC   clone contd. plasmid pJ0101 which contd. a 1.35 kb PstI sequence.
CC   This fragment was isolated, subjected to PCR and the amplification
CC   prod. ligated with the 2.2 kb BglII-SacI fragment of pRU276 (contg.
CC   a synthetic Tac promoter and transcription stop signal).  The
CC   recombinant plasmid was designated pRGLN2.1.
XX
SQ   Sequence 1378 BP; 397 A; 287 C; 276 G; 418 T; 0 other;


  Query Match             38.6%;  Score 489.2;  DB 14;  Length 1378;
  Best Local Similarity   65.0%;  Pred. No. 3.7e-145;
  Matches  722;  Conservative    0;  Mismatches  388;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||     |||||||||||| | || ||     |    || || ||  ||||
Db    136 accctgaagtgaatatgaatattagtcagatgatttcctactggggatacccaagtgaaa 195

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          ||||||||||| |  ||||||||||| |||||||||   |||||||| ||||||   ||
Db    196 aatatgaagttgtgactgaagatggttatatccttgaggttaacagaattccttatggga 255

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |   |   ||| | |||    || || || || ||  |||||||||  |  |||
Db    256 agaaaaattcagggaacagaggccagagacccgtcgtatttttgcagcatggtttgcttg 315

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |   ||    |||||||| |||||||||||||||||| ||||||| ||| ||||||||||
Db    316 catcagcctcaaactggatctccaacctgcccaacaacagcctggcctttattctggcag 375

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          |||||||||| ||  |||||| |||||||||||||| |||||  |||||||  ||| | ||
Db    376 atgctggttatggtgtgtggctggggaacagcagaggaaatacctggtccagaagaaact 435

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                ||   ||||   | ||| |||||||||||| || | |||||| |||||||  |
Db    436 tatactattcaccagactccgttgaattctgggcttttagctttgatgaaatggctaaat 495

Qy    485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
           |||||||| ||    ||  |||| ||| |  || |||| || |||||||||  |  | |
Db    496 atgaccttccagccacaattgacttcattgtaaaggaaactggacaggaaaagctgcact 555

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          |||| ||  |||| ||||||||||||||| || |||||||| |||||  || |  ||  |||
Db    556 atgttggtcattcccagggcaccaccattggttttattgccttttctacaaatcccaagc 615

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          |||||  | | ||||||||  | || ||| |||||  ||  | |||||||| ||| || |||
Db    616 tggctgaaagaatcaaaacctttttatgcattagctccagtcgccactgtgaagtatacaa 675

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||||||  |  | |||| ||  |||  | ||     |  ||| |||||    | |||||
Db    676 aaagccttgtaaacaaacttaggtttattcctccaaccatgttcaagattatatttggtg 735

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||       || |     |||     |   ||||  | | |  | |||  ||
Db    736 acaaaatattctacccacacaatttctttgatcaatttcttgccactcaagtgtgttccc 795

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||  ||  ||   ||||| || |||  | | ||  |  |  |||| ||  ||| |
Db    796 gtgagacactgaatgtaatttgcagcaatgccttatttatcatttgtggctttgacagcg 855

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  |||||||||| ||      ||| || |   | || | ||  || ||||| || |
Db    856 caaacttgaacatgagtcgcttggatgtgtacgtgtcacataatccggcaggaacttcag 915

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | ||||| || || ||||||| |||||| || || |||||  |  | |  || ||| | |
Db    916 ttcaaaacatgctgcactggacccaggctgttaaatctggaaattttcaagcttttaatt 975

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| |||    |  | ||| | |   | | |||||||||| || |||      |||| |
Db    976 ggggaagtccagctcagaatgtagtgcacttcaatcagcccacacctccctactacaatg 1035

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           | |  |  |||| || || |  ||| ||||||  || ||  | ||||| |  |  | |
Db    1036 tgaccgccatgaatgtgccaattgcagtgtggagtggtggcaatgactggttggctgacc 1095

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           |  |||| ||  |  |  ||||  |  |  |  | || ||||| ||||| ||| | ||||
Db    1096 cccaagatgttgacctttgcttccaaaactttctaatctcatttaccacaaggagattc 1155

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
            |  ||    |||| |||||||| |||||||   || ||||||||||     | |||||||
Db    1156 ttccatataatcacttggattttatctgggcaatgaatgctcctcaagaagtttacaatg 1215

Qy    1205 aaatcatccatctgatgcagcaggaggaga 1234
           |||| ||   | |||||    | ||  | |
Db    1216 aaattatttctatgatggcaaaagataaaa 1245


RESULT   5
AAN60685
ID   AAN60685 standard; DNA; 1367 BP.
XX
AC   AAN60685;
XX
DT   08-JUN-1991  (first entry)
XX
DE   Sequence encoding pregastric lipase.
XX
KW   Precursor polypeptide; secretion vector; enzyme; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     47..103
FT                   /*tag=  a
FT   mat_peptide     104..1243
FT                   /*tag=  b
XX
PN   WO8603778-A.
XX
PD   03-JUL-1986.
XX
PF   23-DEC-1985;   85WO-GB00599.
XX
PR   21-DEC-1984;   84GB-0032483.
PR   23-DEC-1985;   85WO-GB00599.
PR   01-JAN-1986;   86GB-0019568.
XX
PA   (BREW-) BREWING RES FOUND.
PA   (TUBB/) TUBB R S.
XX
PI   Tubb RS;
XX
DR   WPI; 1986-182910/28.
DR   P-PSDB; AAP60724.
XX
PT   New precursor polypeptide of defined sequence - and corresp. DNA
PT   used to transform hosts for prodn. of the polypeptide
XX
PS   Example; Fig 5; 60pp; English.
XX
CC   The inventors claim a DNA sequence encoding AAP60723 linked to a
CC   promoter upstream and a gene for a polypeptide downstream.
CC   Particular examples are the yeast emzyme AMG, the mammalian enzyme,
CC   gastric lipase and the mammalian lymphokine, interferon-alpha2.
```

```
XX
SQ   Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;


  Query Match             37.0%;  Score 470;  DB 7;  Length 1367;
  Best Local Similarity   64.2%;  Pred. No. 4.8e-139;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy     119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
           |||    ||| ||||     ||||| |||||| | || ||     ||   || || ||
Db     123 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 182

Qy     179 gtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctc 238
           ||| ||||||||||| |  |||||||||| ||||| |||   || || || ||||||
Db     183 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 242

Qy     239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
             ||       | |   ||| |||||    || ||||| |||||  ||||||||||
Db     243 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 302

Qy     299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
           |  ||| |   || |  ||||||||||||||||||||| ||||| ||||| | ||||||||
Db     303 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 362

Qy     359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
           |||||||||||||| ||| |||||| |||| |||||||| |||||| ||||| |  |||
Db     363 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 422

Qy     419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
           | ||  |      ||   |||   || ||| |||||||||||||| | |||||| ||||
Db     423 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 482

Qy     479 ctaggtttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaaga 538
           |||  | |||||||||| ||    ||  |||| ||| |  |||||||  || ||| |  ||
Db     483 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 542

Qy     539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
           |  | ||||| ||| |||| |||||||||||||| || |||||||  ||||||||||  |
Db     543 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 602

Qy     599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
           |    |||||| | | ||||||||  |  | |||| |||| ||  | ||||||||  |||
Db     603 ccagcctggctaaaagaatcaaaaccttctatgctctagctcctgttgccactgtgaagt 662

Qy     659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
           || |||||||||     | |||| ||   ||  | ||  |     |  |||||    | |
Db     663 atacaaaaagccttataaacaaacttagatttgttcctcaatccctcttcaagtttatat 722

Qy     719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 778
           ||||  | |||  |||       || | |   |||     |   |||| || | |  | |
Db     723 ttggtgacaaaatattctacccacacaacttctttgatcaatttcttgctactgaagtgt 782

Qy     779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
           |   ||  | ||| ||  |||   |||| || |||  | | ||  |  |  |||||||
Db     783 gctcccgtgagatgctgaatctcctttgcagcaatgccttatttataatttgtggatttg 842

Qy     839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
           |||  || ||  | |||| ||| ||      ||| |||    | || | ||  || ||||
Db     843 acagtaagaactttaacacgagtcgcttggatgtgtatctatcacataatccagcaggaa 902

Qy     899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
           | ||||| ||||| ||  | || |||| |||||| || || ||||  || |||  || |
Db     903 cttctgttcaaaacatgttccattggacccaggctgttaagtctgggaaattccaagctt 962

Qy     959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
           |||||||||  ||        | |||  |    | |   |||||  |    |||      |
Db     963 atgactggggaagcccagttcagaataggatgcactatgatcagtcccaacctccctact 1022

Qy    1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
           |||  || | ||  ||||  || || |  ||| ||||||  || ||  |||||  | |
Db    1023 acaatgtgacagccatgaatgtaccaattgcagtgtggaacggtggcaaggacctgttgg 1082

Qy    1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
           |  | ||  |||| ||     |  ||||  |  |  |  |||| || || ||||| |||
Db    1083 ctgacccccaagatgttggccttttgcttccaaaactccccaatcttatttaccacaagg 1142

Qy    1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
           | ||||||   |    |||| |||| || |||||||   |||||| |||||     | |
Db    1143 agattccttttacaatcacttggactttatctgggcaatggatgcccctcaagaagttt 1202

Qy    1199 acaatgaaatcatccatctgat 1220
           ||||||| ||  |   | ||||
Db    1203 acaatgacattgtttctatgat 1224


RESULT   6
AAN60566
ID   AAN60566 standard; DNA; 1367 BP.
XX
AC   AAN60566;
XX
DT   22-AUG-1991  (first entry)
XX
DE   Sequence encoding human pregastric lipase.
XX
KW   Cystic fibrosis therapy; enzyme; lipase deficiency; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             47..103
FT                   /*tag=  a*
FT   mat_peptide     104..1243
FT                   /*tag=  b
XX
PN   WO8601532-A.
XX
PD   13-MAR-1986.
XX
PF   15-AUG-1985;   85WO-GB00364.
XX
PR   21-AUG-1984;   84GB-0021210.
PR   15-AUG-1985;   85WO-GB00364.
PR   01-JAN-1986;   86GB-0008897.
XX
PA   (CELL-) CELLTECH LTD.
PA   (LOWE/) LOWE P A.
XX
PI   Lowe PA;
XX
DR   WPI; 1986-081634/12.
DR   P-PSDB; AAP60658.
XX
PT   New gastric lipase protein, esp. of human origin - for treating
PT   lipase deficiency, and DNA sequences coding for it
XX
PS   Disclosure; Fig 3; 39pp; English.
XX
CC   The inventors claim a pregastric lipase protein and a gene encoding
CC   it. Gastric lipase protein is useful for oral administration to
CC   treat lipase deficiency, e.g. cystic fibrosis or pancreatitis.
XX
SQ   Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;


  Query Match             37.0%;  Score 470;  DB 7;  Length 1367;
  Best Local Similarity   64.2%;  Pred. No. 4.8e-139;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy     119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
           |||    ||| ||||     ||||| |||||| | || ||     ||   || || ||
Db     123 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 182

Qy     179 gtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctc 238
           ||| ||||||||||| |  |||||||||| ||||| |||   || || || ||||||
Db     183 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 242
```

```
Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
Db    243 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 302

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
Db    303 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 362

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
Db    363 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 422

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
Db    423 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 482

Qy    479 ctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaaga 538
Db    483 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 542

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
Db    543 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 602

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
Db    603 ccagcctggctaaaagaatcaaaaccttctatgctctagctcctgttgccactgtgaagt 662

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
Db    663 atacaaaaagccttataaacaaacttagatttgttcctcaatccctcttcaagtttatat 722

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 778
Db    723 ttggtgacaaaatattctacccacacaacttctttgatcaatttcttgctactgaagtgt 782

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
Db    783 gctcccgtgagatgctgaatctcctttgcagcaatgccttatttataatttgtggatttg 842

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
Db    843 acagtaagaactttaacacgagtcgcttggatgtgtatctatcacataatccagcaggaa 902

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
Db    903 cttctgttcaaaacatgttccattggacccaggctgttaagtctgggaaattccaagctt 962

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
Db    963 atgactggggaagcccagttcagaataggatgcactatgatcagtcccaacctccctact 1022

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
Db   1023 acaatgtgacagccatgaatgtaccaattgcagtgtggaacggtggcaaggacctgttgg 1082

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
Db   1083 ctgacccccaagatgttggccttttgcttccaaaactccccaatcttatttaccacaagg 1142

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
Db   1143 agattccttttacaatcacttggactttatctgggcaatggatgcccctcaagaagttt 1202

Qy   1199 acaatgaaatcatccatctgat 1220
Db   1203 acaatgacattgtttctatgat 1224
```

RESULT 7
AAT58916
ID AAT58916 standard; DNA; 1367 BP.

XX
AC AAT58916;
XX
DT 19-AUG-1997 (first entry)
XX
DE Human gastric lipase coding sequence.
XX
KW Duodenal; gastric; lipase; transgenic; plant; recombinant; extract; ss;
KW food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;
KW hydrolysis; trans-esterification; substrate; enzyme; biofuel; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 47..1243
FT /*tag= a
FT /product= gastric lipase
FT sig_peptide 47..103
FT /*tag= b
FT mat_peptide 104..1240
FT /*tag= c
XX
PN WO9633277-A2.
XX
PD 24-OCT-1996.
XX
PF 19-APR-1996; 96WO-FR00606.
XX
PR 20-APR-1995; 95FR-0004754.
XX
PA (BIOC-) BIOCEM SA.
PA (LJOU) INST RECH JOUVEINAL.
XX
PI Baudino S, Benicourt C, Cudrey C, Gruber V, Lenee P;
PI Merot B;
XX
DR WPI; 1996-485783/48.
DR P-PSDB; AAW09383.
XX
PT Recombinant human or canine pre-duodenal lipase prodn. in transgenic
PT plants - useful for facilitating absorption of fat, as bio-catalysts
PT and for prodn. of bio-fuel
XX
PS Claim 3; Fig 4; 130pp; French.
XX
CC This is the nucleotide sequence encoding the human pre-duodenal (i.e.
CC gastric) lipase enzyme. The sequence can be used to generate transgenic
CC plants producing recombinant lipase in an enzymatically active form.
CC Alternatively bases encoding amino acids 20-23 or 20-73 (1-4 or 1-54 of
CC the mature protein, respectively) can be deleted to form the derivatives
CC designated delta-4 or delta-54 respectively. Plants, or their extracts,
CC expressing the lipases or the truncated derivatives, can be used:
CC (a) as pharmaceuticals or food to facilitate absorption of fat, either
CC in healthy subjects or in patients with inadequate levels of
CC gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine
CC pancreatic insufficient, in the elderly or in patients undergoing medical
CC treatment which alters fat adsorption; (b) for performing industrial or
CC agricultural reactions, e.g. in processing of fats or in the dairy
CC industry, for hydrolysis or transesterification reactions, etc., where
CC the plant material may provide both enzyme and substrate. The transgenic
CC plants can also be used for biofuel production.
XX
SQ Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;

Query Match 37.0%; Score 470; DB 17; Length 1367;
Best Local Similarity 64.2%; Pred. No. 4.8e-139;
Matches 707; Conservative 0; Mismatches 395; Indels 0; Gaps 0;

```
Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
Db    123 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 182
```

```
Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctc 238
Db    183 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 242

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
Db    243 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 302

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
Db    303 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 362

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
Db    363 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 422

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
Db    423 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 482

Qy    479 ctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaaga 538
Db    483 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 542

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
Db    543 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 602

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
Db    603 ccagcctggctaaaagaatcaaaaccttctatgctctagctcctgttgccactgtgaagt 662

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
Db    663 atacaaaaagccttataaacaaacttagatttgttcctcaatccctcttcaagtttatat 722

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 778
Db    723 ttggtgacaaaatattctacccacacaacttctttgatcaatttcttgctactgaagtgt 782

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
Db    783 gctcccgtgagatgctgaatctcctttgcagcaatgccttatttataatttgtggatttg 842

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
Db    843 acagtaagaactttaacacgagtcgcttggatgtgtatctatcacataatccagcaggaa 902

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
Db    903 cttctgttcaaaacatgttccattggacccaggctgttaagtctgggaaattccaagctt 962

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
Db    963 atgactggggaagcccagttcagaataggatgcactatgatcagtcccaacctccctact 1022

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
Db   1023 acaatgtgacagccatgaatgtaccaattgcagtgtggaacggtggcaaggacctgttgg 1082

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
Db   1083 ctgacccccaagatgttggccttttgcttccaaaactccccaatcttatttaccacaagg 1142

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
Db   1143 agattccttttacaatcacttggactttatctgggcaatggatgcccctcaagaagttt 1202

Qy   1199 acaatgaaatcatccatctgat 1220
Db   1203 acaatgacattgtttctatgat 1224
```

RESULT 8
AAH57420
ID AAH57420 standard; cDNA; 1695 BP.
XX
AC AAH57420;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human stomach cell specific cDNA sequence SEQ ID NO:260.
XX
KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX
OS Homo sapiens.
XX
PN WO200132927-A2.
XX
PD 10-MAY-2001.
XX
PF 02-NOV-2000; 2000WO-US30396.
XX
PR 04-NOV-1999; 99US-0163508.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Sornasse T, Seilhamer JJ, Watson GA;
XX
DR WPI; 2001-291057/30.
XX
PT New cell and tissue specific polynucleotides useful for diagnosis,
PT prognosis or monitoring of treatments for disorders where the gene is
PT associated with a cancer, immunopathology or neuropathology -
XX
PS Claim 1; Page 189-190; 327pp; English.
XX
CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
CC sequences (I). (I) can have cytostatic, immunomodulatory and
CC neuroprotective activities, and can be used in gene therapy. (I) and
CC proteins (II) encoded by then are used in high throughput screening
CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC mimetics, peptides, proteins, agonists, antagonists, antibodies or
CC their fragments, immunoglobulins, inhibitors, drug compounds and
CC pharmaceutical agents. Expression of (I) in a sample indicates the
CC differentiation of embryonic stem cells into a tissue selected from
CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC tissues. (I) and (II) are used to produce an expression profile that
CC defines a metabolic or developmental process, treatment, condition,
CC disease or disorder. The gene profile can be used for diagnosis,
CC prognosis or monitoring of treatments and for investigating a
CC predisposition to a disorder where the gene is associated with a
CC cancer, immunopathology or neuropathology.
XX
SQ Sequence 1695 BP; 480 A; 356 C; 389 G; 470 T; 0 other;

Query Match 36.9%; Score 468.4; DB 22; Length 1695;
Best Local Similarity 64.1%; Pred. No. 1.8e-138;
Matches 706; Conservative 0; Mismatches 396; Indels 0; Gaps 0;

```
Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
Db    116 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 175

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctc 238
Db    176 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 235

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
Db    236 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 295
```

```
Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
Db    296 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 355

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
Db    356 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 415

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
Db    416 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 475

Qy    479 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 538
Db    476 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaagctggacagaagcagc 535

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
Db    536 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 595

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
Db    596 ccagcctggctaaaagaatcaaaaccttctatgctctagctcctgttgccactgtgaagt 655

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
Db    656 atacaaaaagccttataaacaaacttagatttgttcctcaatccctcttcaagtttatat 715

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 778
Db    716 ttggtgacaaaatattctacccacacaacttctttgatcaatttcttgctactgaagtgt 775

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
Db    776 gctcccgtgagatgctgaatctcctttgcagcaatgccttatttataatttgtggatttg 835

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
Db    836 acagtaagaacttaacacgagtcgcttggatgtgtatctatcacataatccagcaggaa 895

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
Db    896 cttctgttcaaaacatgttccattggacccaggctgttaagtctgggaaattccaagctt 955

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
Db    956 atgactggggaagcccagttcagaataggatgcactatgatcagtcccaacctccctact 1015

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
Db   1016 acaatgtgacagccatgaatgtaccaattgcagtgtggaacggtggcaaggacctgttgg 1075

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
Db   1076 ctgacccccaagatgttggccttttgcttccaaaactccccaatcttatttaccacaagg 1135

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
Db   1136 agattccttttacaatcacttggactttatctgggcaatggatgcccctcaagaagttt 1195

Qy   1199 acaatgaaatcatccatctgat 1220
Db   1196 acaatgacattgtttctatgat 1217
```

RESULT 9
AAN50385
ID AAN50385 standard; DNA; 1336 BP.
XX
AC AAN50385;
XX
DT 17-JAN-1992 (first entry)
XX
DE Rat prelingual lipase gene.
XX
KW Prelingual lipase; enzyme; EC-3.1.1.3; ss.
XX
OS Rattus rattus.
XX
FH Key Location/Qualifiers
FT CDS 7..1194
FT /*tag= a
XX
PN GB2142337-A.
XX
PD 16-JAN-1985.
XX
PF 29-JUN-1984; 84GB-0016581.
XX
PR 01-JUL-1983; 83GB-0017989.
PR 05-SEP-1983; 83GB-0023759.
XX
PA (CELL-) CELLTECH LTD.
XX
PI Carey NH, Williamson R;
XX
DR WPI; 1985-014450/03.
DR P-PSDB; AAP50322.
XX
PT New lingual lipase protein for treatment of lipase deficiency - also
PT new pre-lingual lipase protein and related products
XX
PS Disclosure; Fig 5; 15pp; English.
XX
CC This sequence may be expressed in a transformant host organism using
CC a vector plasmid in order to produce a prelingual lipase protein.
CC The expressed protein may be used for the treatment of lipase
CC deficiency.
XX
SQ Sequence 1336 BP; 409 A; 285 C; 263 G; 379 T; 0 other;

Query Match 36.6%; Score 464.4; DB 6; Length 1336;
Best Local Similarity 63.4%; Pred. No. 2.9e-137;
Matches 711; Conservative 0; Mismatches 411; Indels 0; Gaps 0;

```
Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
Db     80 ctggaaaccctgaagcaaatatgaatattagtcagatgataacttactggggatatccat 139

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
Db    140 gtcaagaatatgaagttgttactgaagatggctacattctgggggtctacagaattcctc 199

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
Db    200 atgggaagaataattctgaaaatataggcaagagacctgtggtgtattgcagcatggtt 259

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
Db    260 tgatcgcatcagccacaaactggattgcaaatctaccaaacaacagcctcgcctttatgc 319

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
Db    320 tagcagacgctggctatgatgtgtggctggggaacagtcgaggaaatacatggtccagga 379

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
Db    380 aaaatgtatactactcaccagactcagttgaattctgggctttcagctttgatgaaatgg 439

Qy    479 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 538
Db    440 ctaaatatgaccttcccgccacaataaacttcattgtacagaaaactggacaagagaaga 499

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
```

```
Db    500 tacactatgttggtcattctcagcgcaccactattggtttcattgccttttct                 559

Qy    599 cagagctggctcagcaaaatcaaaatgtatttttgcttagcacccatagccactgttaagc 658

Db    560 ctacactggccaaaaaaatcaagacgtttttatgcattagctccagttgctaccgtgaagt 619

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718

Db    620 atacacaaagtcccttgaaaaagatttcatttattcctacatttctttcaagcttatgt 679

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 778

Db    680 ttggcaagaaaatgttcctgccccacacctactttgatgactttcttggtaccgaagtgt 739

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838

Db    740 gctcacgggaggttctagatcttctctgcagcaacactttattcatcttctgtggatttg 799

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898

Db    800 acaagaaaaacttaaatgtgagtcgttttgatgtgtatctagggcataatccagcaggga 859

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958

Db    860 catctgttcaagactttctccactgggcacagcttgttagatctgggaaatttcaagcct 919

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018

Db    920 tcaactggggaagcccatcccagaacatgttacactacaaccagaaaacgcctcctgaat 979

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078

Db    980 atgatgtgtcagccatgactgtgccagttgcagtgtggaacggtggcaatgacatcctgg 1039

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138

Db   1040 ctgatccccaagatgtcgccatgctgcttcccaaactctccaacctcctgttccataagg 1099

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198

Db   1100 agattcttgcctacaatcacctggacttcatctgggcaatggatgcccctcaagaggttt 1159

Qy   1199 acaatgaaatcatccatctgatgcagcaggaggagaccaacc 1240

Db   1160 acaatgagatgatttccatgatggcagaagactaaaagaatc 1201


RESULT 10
AAQ68388
ID   AAQ68388 standard; DNA; 1528 BP.
XX
AC   AAQ68388;
XX
DT   20-FEB-1995  (first entry)
XX
DE   Canine gastric lipase.
XX
KW   Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
KW   bioconversion; exocrine pancreatic insufficiency; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1140
FT                   /*tag=  a
FT   misc_RNA        1..1137
FT                   /*tag=  b
FT                   /note= "claim 1"
XX
PN   WO9413816-A.
XX
PD   23-JUN-1994.
XX
PF   16-DEC-1993;  93WO-FR01260.
XX
PR   16-DEC-1992;  92FR-0015201.
XX
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Benicourt C,  Blanchard C,  Junien J;
XX
DR   WPI; 1994-217890/26.
DR   P-PSDB; AAR56870.
XX
PT   Recombinant canine gastric lipase and nucleic acid encoding it -
PT   are used for improving absorption of ingested fat, treating
PT   mucoviscidosis etc. and in enzymatic bio-conversions
XX
PS   Claim 1; Fig 8; 52pp; French.
XX
CC   The sequence given below is the sequence of figure 8, altered
CC   according to the amendments described on page 2 of the appended
CC   letter.
CC   CGL is used to improve absorption of ingested fat, in healthy and
CC   sick patients (e.g. having altered levels of gastric lipase); to
CC   treat conditions associated with insufficiency (or lack) of lipases,
CC   esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC   where immobilised, for bioconversions, e.g. hydrolysis or
CC   transesterification (other mammalian gastric lipases, or derivs.,
CC   can be used in this application).
XX
SQ   Sequence 1528 BP; 394 A; 386 C; 328 G; 420 T; 0 other;


  Query Match             35.0%;  Score 444.2;  DB 15;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 8.3e-131;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184

Db     26 accctgaagtgaccatgaatataagtcagatgatcacctactggggatacccagctgagg 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttctgttaacaggattcctcgaggcc 244

Db     86 aatatgaagttgtgaccgaagacggttatatccttgggatcgacagaattccttatggga 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304

Db    146 ggaaaaattcagagaatataggccggagacctgttgcatttttgcaacacggtttgctcg 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364

Db    206 catcagccacaaactggatctccaacctgcccaacaacagcctggccttcatcctggccg 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424

Db    266 acgccgggtacgacgtgtggctggggaacagcaggggcaacacctgggccaggaggaatc 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484

Db    326 tgtactactcgcccgactccgtcgaattctgggctttcagctttgacgagatggctaaat 385

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544

Db    386 atgaccttcccgccaccattgacttcatcttgaagaaaacgggacaggacaagctacact 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604

Db    446 acgttggccattcccagggcaccaccattggtttcatcgccttttccaccaatcccaagc 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664

Db    506 tggcgaaacggatcaaaaccttctatgcattagctcccgttgccaccgtgaagtacaccg 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
```

```
Db    566 aaaccctgttaaacaaactcatgctcgtcccttcgttcctcttcaagcttatatttggaa 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784

Db    626 acaaaatattctacccacaccacttctttgatcaatttctcgccaccgaggtatgctccc 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844

Db    686 gcgagacggtggatctcctctgcagcaacgccctgtttatcatttgtggatttgacacta 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904

Db    746 tgaacttgaacatgagtcgcttggatgtgtatctgtcacataatccagcaggaacatcgg 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964

Db    806 ttcagaacgtgctccactggtcccaggctgttaagtctgggaagttccaagcttttgact 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024

Db    866 ggggaagcccagttcagaacatgatgcactatcatcagagcatgcctccctactacaacc 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084

Db    926 tgacagacatgcatgtgccaatcgcagtgtggaacggtggcaacgacttgctggccgacc 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144

Db    986 ctcacgatgttgacctttgctttccaagctccccaatctcatttaccacaggaagattc 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204

Db   1046 ctccttacaatcacttggactttatctgggccatggatgcccctcaagcggtttacaatg 1105

Qy   1205 aaatcatccatctgatg 1221

Db   1106 aaattgtttccatgatg 1122
```

```
RESULT  11
AAT58915
ID   AAT58915 standard; cDNA to mRNA; 1528 BP.
XX
AC   AAT58915;
XX
DT   19-AUG-1997  (first entry)
XX
DE   Dog gastric lipase coding sequence.
XX
KW   Dog; duodenal; gastric; lipase; transgenic; plant; recombinant; extract;
KW   food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;
KW   hydrolysis; trans-esterification; substrate; enzyme; biofuel; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1140
FT                   /*tag= a
FT                   /product= gastric lipase
XX
PN   WO9633277-A2.
XX
PD   24-OCT-1996.
XX
PF   19-APR-1996;   96WO-FR00606.
XX
PR   20-APR-1995;   95FR-0004754.
XX
PA   (BIOC-) BIOCEM SA.
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Baudino S,  Benicourt C,  Cudrey C,  Gruber V,  Lenee P;
PI   Merot B;
XX
DR   WPI; 1996-485783/48.
DR   P-PSDB; AAW09382.
XX
PT   Recombinant human or canine pre-duodenal lipase prodn. in transgenic
PT   plants - useful for facilitating absorption of fat, as bio-catalysts
PT   and for prodn. of bio-fuel
XX
PS   Claim 2; Fig 1; 130pp; French.
XX
CC   This is the nucleotide sequence encoding the dog pre-duodenal (i.e.
CC   gastric) lipase enzyme.  The sequence can be used to generate transgenic
CC   plants producing recombinant lipase in an enzymatically active form.
CC   Alternatively the first 12 bases of this sequence (encoding amino acids
CC   1-4 of the protein encoded by this sequence) or bases 1-162 of this
CC   sequence (encoding amino acids 1-54) can be deleted to form the
CC   derivatives designated delta-4 or delta-54 respectively.  Plants, or
CC   their extracts, expressing the lipases or the truncated derivatives, can
CC   be used: (a) as pharmaceuticals or food to facilitate absorption of fat,
CC   either in healthy subjects or in patients with inadequate levels of
CC   gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine
CC   pancreatic insufficient, in the elderly or in patients undergoing medical
CC   treatment which alters fat adsorption; (b) for performing industrial or
CC   agricultural reactions, e.g. in processing of fats or in the dairy
CC   industry, for hydrolysis or transesterification reactions, etc., where
CC   the plant material may provide both enzyme and substrate.  The transgenic
CC   plants can also be used for biofuel production.
XX
SQ   Sequence 1528 BP; 392 A; 387 C; 329 G; 420 T; 0 other;
```

```
  Query Match             35.0%;  Score 444.2;  DB 17;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 8.3e-131;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184

Db     26 accctgaagtgaccatgaatataagtcagatgatcacctactggggatacccagctgagg 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244

Db     86 aatatgaagttgtgaccgaagacggttatatccttgggatcgacagaattccttatggga 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304

Db    146 ggaaaaattcagagaatataggccggagacctgttgcatttttgcaacacggtttgctcg 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364

Db    206 catcagccacaaactggatctccaacctgcccaacaacagcctggccttcatcctggccg 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424

Db    266 acgccgggtacgacgtgtggctggggaacagcaggggcaacacctgggccaggaggaatc 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484

Db    326 tgtactactcgcccgactccgtcgaattctgggctttcagctttgacgagatggctaaat 385

Qy    485 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 544

Db    386 atgaccttcccgccaccattgacttcatcttgaagaaaacgggacaggacaagctacact 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604

Db    446 acgttggccattcccagggcaccaccattggtttcatcgccttttccaccaatcccaagc 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664

Db    506 tggcgaaacggatcaaaaccttctatgcattagctcccgttgccaccgtgaagtacaccg 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
```

```
Db    566 aaaccctgttaaacaaactcatgctcgtcccttcgttcctcttcaagcttatatttggaa 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||       ||   |   |||     |   || |  |   |  | ||   ||
Db    626 acaaaatattctacccacaccacttctttgatcaatttctcgccaccgaggtatgctccc 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||   | ||||   | || || ||   | ||||  |  |  ||||||  |||| |
Db    686 gcgagacggtggatctcctctgcagcaacgccctgtttatcatttgtggatttgacacta 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  ||||||||||| ||      ||| |||    | || | ||  || ||||||||
Db    746 tgaacttgaacatgagtcgcttggatgtgtatctgtcacataatccagcaggaacatcgg 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | ||  ||  | ||  ||||||  ||||||  || ||  |||||  |  |||  || |||||||
Db    806 ttcagaacgtgctccactggtcccaggctgttaagtctgggaagttccaagcttttgact 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          ||||  ||        | ||  ||    | |   |||||   |  |||      ||||
Db    866 ggggaagcccagttcagaacatgatgcactatcatcagagcatgcctccctactacaacc 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   ||  || |  ||| ||||||  ||  ||  | |||| |||  |  | |
Db    926 tgacagacatgcatgtgccaatcgcagtgtggaacggtggcaacgacttgctggccgacc 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  | ||  ||  |  |  |||| ||  || |  |||| |||| |||| | ||| ||||
Db    986 ctcacgatgttgacctttgctttccaagctccccaatctcatttaccacaggaagattc 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
          ||   |    |||| |||| || |||||||   ||||||| |||||     | |||||||
Db   1046 ctccttacaatcacttggactttatctgggccatggatgcccctcaagcggtttacaatg 1105

Qy   1205 aaatcatccatctgatg 1221
          ||||  |    |||||
Db   1106 aaattgtttccatgatg 1122


RESULT 12
AAQ68389
ID   AAQ68389 standard; DNA; 1531 BP.
XX
AC   AAQ68389;
XX
DT   20-FEB-1995 (first entry)
XX
DE   Canine gastric lipase (Met-CGL).
XX
KW   Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
KW   bioconversion; exocrine pancreatic insufficiency; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1143
FT                   /*tag= a
XX
PN   WO9413816-A.
XX
PD   23-JUN-1994.
XX
PF   16-DEC-1993;   93WO-FR01260.
XX
PR   16-DEC-1992;   92FR-0015201.
XX
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Benicourt C,  Blanchard C,  Junien J;
XX
DR   WPI; 1994-217890/26.
DR   P-PSDB; AAR56871.
XX
PT   Recombinant canine gastric lipase and nucleic acid encoding it -
PT   are used for improving absorption of ingested fat, treating
PT   mucoviscidosis etc. and in enzymatic bio-conversions
XX
PS   Claim 2; Fig 8; 52pp; French.
XX
CC   The sequence given below is the sequence of figure 8, altered
CC   according to the amendments described on page 2 of the appended
CC   letter.
CC   CGL is used to improve absorption of ingested fat, in healthy and
CC   sick patients (e.g. having altered levels of gastric lipase); to
CC   treat conditions associated with insufficiency (or lack) of lipases,
CC   esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC   where immobilised, for bioconversions, e.g. hydrolysis or
CC   transesterification (other mammalian gastric lipases, or derivs.,
CC   can be used in this application).
XX
SQ   Sequence 1531 BP; 395 A; 386 C; 329 G; 421 T; 0 other;


  Query Match             35.0%;  Score 444.2;  DB 15;  Length 1531;
  Best Local Similarity   62.8%;  Pred. No. 8.3e-131;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    ||||||||| ||| | || |||    |    || || ||   |||||
Db     29 accctgaagtgaccatgaatataagtcagatgatcacctactggggatacccagctgagg 88

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          |||||||||| |  || ||||| || |||||||||    |  |||| ||||||   ||
Db     89 aatatgaagttgtgaccgaagacggttatatccttgggatcgacagaattccttatggga 148

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  |  |||| | |||    || |||| |  ||  |||| || ||  |  | |
Db    149 ggaaaaattcagagaatataggccggagacctgttgcatttttgcaacacggtttgctcg 208

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
            |  || |  |||||||| |||||||||||||||| ||||||| |||||| ||||| |
Db    209 catcagccacaaactggatctccaacctgcccaacaacagcctggccttcatcctggccg 268

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||||| ||||||||||||||||| ||| ||||| |  | |   |
Db    269 acgccgggtacgacgtgtggctggggaacagcaggggcaacacctgggccaggaggaatc 328

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
            |     |||    |||   |  || |||||||||||||| | ||| ||||||||||  |
Db    329 tgtactactcgcccgactccgtcgaattctgggctttcagcttgacgagatggctaaat 388

Qy    485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
            |||||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    389 atgaccttcccgccaccattgacttcatcttgaagaaaacgggacaggacaagctacact 448

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          | || ||| |||| |||||||||||||| || || || || |||||||||| ||  |||
Db    449 acgttggccattcccagggcaccaccattggtttcatcgccttttccaccaatcccaagc 508

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||  |    |||||||  |  | ||| |||| ||| | ||||| || ||| |  |
Db    509 tggcgaaacggatcaaaaccttctatgcattagctcccgttgccaccgtgaagtacaccg 568

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||     | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db    569 aaaccctgttaaacaaactcatgctcgtcccttcgttcctcttcaagcttatatttggaa 628

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||       ||   |   |||     |   || |  |   |  | ||   ||
Db    629 acaaaatattctacccacaccacttctttgatcaatttctcgccaccgaggtatgctccc 688

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||   | ||||   | || || ||   | ||||  |  |  ||||||  |||| |
Db    689 gcgagacggtggatctcctctgcagcaacgccctgtttatcatttgtggatttgacacta 748

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
```

```
Db    749 tgaacttgaacatgagtcgcttggatgtgtatctgtcacataatccagcaggaacatcgg 808

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db    809 ttcagaacgtgctccactggtcccaggctgttaagtctgggaagttccaagcttttgact 868

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    869 ggggaagcccagttcagaacatgatgcactatcatcagagcatgcctccctactacaacc 928

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db    929 tgacagacatgcatgtgccaatcgcagtgtggaacggtggcaacgacttgctggccgacc 988

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db    989 ctcacgatgttgacctttgctttccaagctcccaatctcatttaccacaggaagattc 1048

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1049 ctccttacaatcacttggactttatctgggccatggatgcccctcaagcggtttacaatg 1108

Qy   1205 aaatcatccatctgatg 1221
Db   1109 aaattgtttccatgatg 1125
```

RESULT 13
AAF28679
ID AAF28679 standard; cDNA; 699 BP.
XX
AC AAF28679;
XX
DT 05-APR-2001 (first entry)
XX
DE Human protein HP03372 coding sequence #1.
XX
KW Human; hydrophobic domain; immune deficiency; autoimmune disorder;
KW allergy; tissue growth; regeneration; wound healing; burn; tumour;
KW periodontal disease; thrombolytic condition; haemostatic condition;
KW infection; ss.
XX
OS Homo sapiens.
XX
PN WO200102563-A2.
XX
PD 11-JAN-2001.
XX
PF 16-JUN-2000; 2000WO-JP03943.
XX
PR 02-JUL-1999; 99JP-0188835.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
DR WPI; 2001-071581/08.
DR P-PSDB; AAB61608.
XX
PT New human proteins with hydrophobic domains, useful for the treatment
PT of immune disorders, tumors, allergic conditions, thrombosis and
PT microbial infection -
XX
PS Claim 3; Page 113; 153pp; English.
XX
CC The present invention relates to human proteins (AAB61608-AAB61617) and
CC their coding sequences (AAF28679-AAF28698). The proteins of the present
CC invention have hydrophobic domains and can be used for the treatment of
CC various immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, graft-versus-host disease and
CC Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC of allergic reactions and conditions, such as asthma and in regulation of
CC haematopoiesis or lymphoid cell deficiencies. The proteins may also have
CC utility in compositions used for bone, cartilage, tendon and/or nerve
CC tissue growth or regeneration as well as wound healing and in the
CC treatment of burns. The proteins may be used in the treatment of
CC periodontal disease and in other tooth repair processes. Other uses
CC include treatment of thrombolytic and haemostatic conditions, treatment
CC or prevention of tumours and inhibiting infection by bacteria, viruses,
CC fungi and other parasites.
XX
SQ Sequence 699 BP; 216 A; 144 C; 148 G; 191 T; 0 other;

Query Match 22.5%; Score 286; DB 22; Length 699;
Best Local Similarity 67.4%; Pred. No. 1.2e-80;
Matches 403; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

```
Qy     86 gaaatgtgaattcagtacatatgccaactaaagctgtggacccagaagcattcatgaata 145
Db     44 gatctatgtatggttatgacaagaaaggaaacaatgcaaaccctgaagctaatatgaata 103

Qy    146 ttagtgaaatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaag 205
Db    104 ttagccagattatttcttactggggttatccttatgaagagtatgatgttacaacaaaag 163

Qy    206 atgggtatatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacag 265
Db    164 atggttatatccttggaatttataggattccacatggaagaggatgcccagggaggacag 223

Qy    266 gttccaggcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggattt 325
Db    224 ctccaaagcctgctgtgtatttgcagcatggcttaattgcatctgccagtaactggattt 283

Qy    326 ccaacctgcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtgga 385
Db    284 gcaacctgcccaacaacagtttggctttccttctggcagatagtggttatgacgtgtggt 343

Qy    386 tggggaacagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaag 445
Db    344 tggggaacagccgaggaaacacttggtccagaaaacaccttaaattgtcaccgaaatcac 403

Qy    446 atgagttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataa 505
Db    404 cagaatactgggccttcagtttggatgagatggctaaatatgaccttccagccacaatca 463

Qy    506 actttattttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggca 565
Db    464 attttatcatagagaaaactggacagaagcgactctactacgtgggccactcacaaggca 523

Qy    566 ccaccatgggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgt 625
Db    524 ccaccatagcttttatagcattttctacaaacccagaactggctaaaaagattaagatat 583

Qy    626 attttgctttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatt 683
Db    584 tttttgcactggctccagttgtcacagttaaatacacccaaagtcctatgaaaaaact 641
```

RESULT 14
AAF28689
ID AAF28689 standard; cDNA; 1308 BP.
XX
AC AAF28689;
XX
DT 05-APR-2001 (first entry)
XX
DE Human protein HP03372 coding sequence #2.
XX
KW Human; hydrophobic domain; immune deficiency; autoimmune disorder;
KW allergy; tissue growth; regeneration; wound healing; burn; tumour;
KW periodontal disease; thrombolytic condition; haemostatic condition;
KW infection; ss.

XX
OS Homo sapiens.
XX
PN WO200102563-A2.
XX
PD 11-JAN-2001.
XX
PF 16-JUN-2000; 2000WO-JP03943.
XX
PR 02-JUL-1999; 99JP-0188835.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
DR WPI; 2001-071581/08.
DR P-PSDB; AAB61608.
XX
PT New human proteins with hydrophobic domains, useful for the treatment
PT of immune disorders, tumors, allergic conditions, thrombosis and
PT microbial infection -
XX
PS Claim 4; Pages 122-124; 153pp; English.
XX
CC The present invention relates to human proteins (AAB61608-AAB61617) and
CC their coding sequences (AAF28679-AAF28698). The proteins of the present
CC invention have hydrophobic domains and can be used for the treatment of
CC various immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, graft-versus-host disease and
CC Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC of allergic reactions and conditions, such as asthma and in regulation of
CC haematopoiesis or lymphoid cell deficiencies. The proteins may also have
CC utility in compositions used for bone, cartilage, tendon and/or nerve
CC tissue growth or regeneration as well as wound healing and in the
CC treatment of burns. The proteins may be used in the treatment of
CC periodontal disease and in other tooth repair processes. Other uses
CC include treatment of thrombolytic and haemostatic conditions, treatment
CC or prevention of tumours and inhibiting infection by bacteria, viruses,
CC fungi and other parasites.
XX
SQ Sequence 1308 BP; 402 A; 246 C; 264 G; 396 T; 0 other;

Query Match 22.5%; Score 286; DB 22; Length 1308;
Best Local Similarity 67.4%; Pred. No. 1.8e-80;
Matches 403; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

```
Qy      86 gaaatgtgaattcagtacatatgccaactaaagctgtggacccagaagcattcatgaata 145
           ||  | || ||       | | |  |   ||   ||   |||| |||||    |||||||
Db     119 gatctatgtatggttatgacaagaaaggaaacaatgcaaaccctgaagctaatatgaata 178

Qy     146 ttagtgaaatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaag 205
           ||||  | || ||     |    || ||||| | ||| || ||||| ||  ||||  |||
Db     179 ttagccagattatttcttactggggttatccttatgaagagtatgatgttacaacaaaag 238

Qy     206 atgggtatatccttttctgttaacaggattcctcgaggcctagtgcaacctaagaagacag 265
           |||| |||||||||    || | |||||||| |  ||   ||     ||   || |||||
Db     239 atggttatatccttggaatttataggattccacatggaagaggatgcccagggaggacag 298

Qy     266 gttccaggcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggattt 325
           | | | |||||  ||||   |||||||||| || ||| |  ||| || ||||||||||
Db     299 ctccaaagcctgctgtgtatttgcagcatggcttaattgcatctgccagtaactggattt 358

Qy     326 ccaacctgcccaacaatagcctgggcttcattctggcagatgctggtttttgacgtgtgga 385
           ||||||||||||||| ||  |||  ||| |||||||||||  |||||  ||||||||||
Db     359 gcaacctgcccaacaacagtttggctttccttctggcagatagtggttatgacgtgtggt 418

Qy     386 tggggaacagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaag 445
           |||||||||||  | ||||||  | |||||  ||||||||   | | | ||     |   |
Db     419 tggggaacagccgaggaaacacttggtccagaaaacaccttaaattgtcaccgaaatcac 478

Qy     446 atgagttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataa 505
             || | |||||| |||||||  ||||||||||||  | |||||||||| ||    || |
Db     479 cagaatactgggccttcagtttggatgagatggctaaatatgaccttccagccacaatca 538

Qy     506 actttatttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggca 565
           | |||||  |  |||||||| || ||| |     |||| || || ||| | ||||| ||||
Db     539 attttatcatagagaaaactggacagaagcgactctactacgtgggccactcacaaggca 598

Qy     566 ccaccatgggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgt 625
           |||||||  |  ||||| |||||||| || |  ||||| |||||| | || || || || |
Db     599 ccaccatagcttttatagcattttctacaaacccagaactggctaaaaagattaagatat 658

Qy     626 attttgctttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatt 683
           ||||||  | || ||  | | ||| |||||  |  |  |||| ||   ||  ||| |
Db     659 tttttgcactggctccagttgtcacagttaaatacacccaaagtcctatgaaaaaact 716
```

RESULT 15
AAA44349
ID AAA44349 standard; cDNA; 617 BP.
XX
AC AAA44349;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:924.
XX
KW Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200021991-A1.
XX
PD 20-APR-2000.
XX
PF 15-OCT-1999; 99WO-US24206.
XX
PR 15-OCT-1998; 98US-0104436.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR;
XX
DR WPI; 2000-317938/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (sESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 441; 803pp; English.
XX
CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC tissue sources. The sESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;

```
CC   nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC   anticonvulsant; and antidepressant. The sESTs can be used for gene
CC   therapy and in vaccines. The sESTs are useful as probes for the
CC   identification and isolation of full-length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporosis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC   disease), tumours, bacterial, fungal or viral infections, depression and
CC   psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 617 BP; 181 A; 125 C; 139 G; 172 T; 0 other;


  Query Match             20.3%;  Score 257.8;  DB 21;  Length 617;
  Best Local Similarity   70.7%;  Pred. No. 1.1e-71;
  Matches  343;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
          |||    ||| ||||     ||||| |||||| | || ||     ||   || || ||
Db    133 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 192

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
           ||| ||||||||||| |  ||||||||||| ||||| |||   || || || ||||||
Db    193 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 252

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
            ||        |  |   ||| |||||    || ||||| |||||  ||||||||||
Db    253 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 312

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
          |  ||| |   || |  ||||||||||||||||||| ||||| ||||| | ||||||||
Db    313 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 372

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
          |||||||||||||||| ||| |||||| |||| |||||||| |||||| ||||| |  |||
Db    373 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 432

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
           | ||  |      ||   |||   || ||| |||||||||||||| | |||||| ||||
Db    433 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 492

Qy    479 ctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaaga 538
          |||  | ||||||||| ||    ||  |||| ||| |  ||||||| || ||| |  ||
Db    493 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 552

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
          |  | ||||| ||| |||| |||||||||||||| || |||||||| ||||||||||  |
Db    553 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 612

Qy    599 cagag 603
           |||
Db    613 tcgag 617


Search completed: January 30, 2002, 22:13:51
Job time: 3898 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 30, 2002, 20:27:03 ; Search time 90.11 Seconds
(without alignments)
3189.442 Million cell updates/sec

Title: US-09-578-063-46
Perfect score: 1269
Sequence: 1 atgttggaaaccttgtcaag.....gacggtgtgaggccgtattg 1269

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	444.2	35.0	1137	1	US-08-227-108-2	Sequence 2, Appli
	2	444.2	35.0	1137	2	US-09-073-674-2	Sequence 2, Appli
	3	444.2	35.0	1140	1	US-08-227-108-4	Sequence 4, Appli
	4	444.2	35.0	1140	2	US-09-073-674-4	Sequence 4, Appli
	5	444.2	35.0	1146	1	US-08-227-108-6	Sequence 6, Appli
	6	444.2	35.0	1146	2	US-09-073-674-6	Sequence 6, Appli
	7	444.2	35.0	1528	1	US-08-227-108-1	Sequence 1, Appli
	8	444.2	35.0	1528	2	US-09-073-674-1	Sequence 1, Appli
	9	48.2	3.8	178	1	US-08-751-782-2	Sequence 2, Appli
	10	48.2	3.8	178	2	US-08-925-171-2	Sequence 2, Appli
	11	46.4	3.7	43676	3	US-09-356-952-12	Sequence 12, Appl
	12	36.6	2.9	3271	2	US-08-852-806-1	Sequence 1, Appli
	13	36.6	2.9	3271	3	US-09-163-669-1	Sequence 1, Appli
	14	36.6	2.9	3282	1	US-08-276-852-154	Sequence 154, App
c	15	36.6	2.9	3282	1	US-08-276-852-169	Sequence 169, App
	16	36.6	2.9	3282	1	US-08-899-575-154	Sequence 154, App
c	17	36.6	2.9	3282	1	US-08-899-575-169	Sequence 169, App
	18	36.6	2.9	3282	1	US-08-899-575-154	Sequence 154, App
c	19	36.6	2.9	3282	1	US-08-899-575-169	Sequence 169, App
	20	36.6	2.9	3282	5	PCT-US95-08743-154	Sequence 154, App
c	21	36.6	2.9	3282	5	PCT-US95-08743-169	Sequence 169, App
	22	36.6	2.9	13254	1	US-08-276-852-156	Sequence 156, App
c	23	36.6	2.9	13254	1	US-08-276-852-170	Sequence 170, App
	24	36.6	2.9	13254	1	US-08-899-575-156	Sequence 156, App
c	25	36.6	2.9	13254	1	US-08-899-575-170	Sequence 170, App
	26	36.6	2.9	13254	1	US-08-899-575-156	Sequence 156, App
c	27	36.6	2.9	13254	1	US-08-899-575-170	Sequence 170, App
	28	36.6	2.9	13254	5	PCT-US95-08743-156	Sequence 156, App
c	29	36.6	2.9	13254	5	PCT-US95-08743-170	Sequence 170, App
	30	33.6	2.6	8392	1	US-08-080-255-6	Sequence 6, Appli
	31	33.6	2.6	8392	3	US-08-465-713-6	Sequence 6, Appli
	32	33.6	2.6	8392	5	PCT-US93-05857-6	Sequence 6, Appli
c	33	32.6	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	34	32	2.5	2197	1	US-08-233-005-3	Sequence 3, Appli
c	35	32	2.5	2197	1	US-08-428-943-3	Sequence 3, Appli
c	36	32	2.5	2197	3	US-09-016-649-3	Sequence 3, Appli
c	37	32	2.5	2197	5	PCT-US95-04858-3	Sequence 3, Appli
	38	32	2.5	8342	3	US-08-545-860D-63	Sequence 63, Appl
	39	32	2.5	8342	5	PCT-US94-04496-63	Sequence 63, Appl
	40	31.6	2.5	3343	4	US-08-965-762-19	Sequence 19, Appl
	41	31	2.4	3212	3	US-08-673-814-1	Sequence 1, Appli
c	42	30.4	2.4	1529	1	US-08-336-778-1	Sequence 1, Appli
c	43	30.2	2.4	4982	3	US-08-699-103B-1	Sequence 1, Appli
	44	30.2	2.4	10079	2	US-08-476-866-20	Sequence 20, Appl
c	45	30	2.4	3639	2	US-08-737-524B-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-08-227-108-2
; Sequence 2, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
US-08-227-108-2

Query Match 35.0%; Score 444.2; DB 1; Length 1137;
Best Local Similarity 62.8%; Pred. No. 1.5e-134;
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    ||||||||| ||| | || |||    |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          |||||||||| |  || ||||| || |||||||||    |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
             |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||| |||||||||||||||| ||||||| |||||| |||||
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||| |||||||||||||||| ||| ||||| |  | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
          |     |||    |||   |  || |||||||||||||| | ||| ||||||||||  |
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||||  ||    ||  |||| || ||| ||||||||||  ||||| ||| |  | |
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||     | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||       ||   |  |||      |  || | |  |  | | ||   ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
          | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            || |||||||||| ||      ||| |||   | || | ||  || ||||||||| |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | || ||  | || ||||||  |||||| || || |||||  |  |||  || |||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          ||||  ||       | ||  ||    | |   |||||   |  |||      ||||
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
          ||   |    |||| |||| || |||||||   |||||||  |||||     | |||||||
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105
```

```
Qy   1205 aaatcatccatctgatg 1221
          ||||  |     |||||
Db   1106 AAATTGTTTCCATGATG 1122
```

RESULT 2
US-09-073-674-2
; Sequence 2, Application US/09073674
; Patent No. 5998189
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warner-Lambert Company
; STREET: 2800 Plymouth Road
; CITY: Ann Arbor
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,674
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crissey, Todd M.
; REGISTRATION NUMBER: 37,807
; REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 734 622-7530
; TELEFAX: 734 622-1553
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
US-09-073-674-2

Query Match 35.0%; Score 444.2; DB 2; Length 1137;
Best Local Similarity 62.8%; Pred. No. 1.5e-134;
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    ||||||||| ||| | || |||    |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          |||||||||| |  || ||||| || |||||||||    |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
             |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||| |||||||||||||||| ||||||| |||||| |||||
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
```

```
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaactttatttgcagaaaacgggccaggaaaagatctatt 544
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtatttttgctttagcacccatagccactgttaagcatgcaa 664
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaattttgttgctgccagatatgatgatcaagggattgtttggca 724
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1205 aaatcatccatctgatg 1221
Db   1106 AAATTGTTTCCATGATG 1122

RESULT    3
US-08-227-108-4
; Sequence 4, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1140
US-08-227-108-4

  Query Match             35.0%;  Score 444.2;  DB 1;  Length 1140;
  Best Local Similarity   62.8%;  Pred. No. 1.5e-134;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db     29 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 88

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttctgttaacaggattcctcgaggcc 244
Db     89 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 148

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    149 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 208

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    209 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 268

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db    269 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 328

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    329 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 388

Qy    485 ttgaccttcctgcagtgataaactttatttgcagaaaacgggccaggaaaagatctatt 544
Db    389 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 448

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db    449 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 508

Qy    605 tggctcagaaaatcaaaatgtatttttgctttagcacccatagccactgttaagcatgcaa 664
Db    509 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 568
```

```
Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||     | |||| |  || |  | ||      |  |  |||||     | |||||
Db    569 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 628

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||       ||  .  |   |||     |   || |  |   |  | ||   ||
Db    629 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 688

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    689 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 748

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  ||||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    749 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 808

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | || ||  | || ||||||  |||||| || || |||||  |  |||  || |||||||
Db    809 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 868

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| ||       | ||  ||    | |   |||||   |  |||      ||||
Db    869 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 928

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db    929 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 988

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    989 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1048

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
          ||   |   |||| |||| || |||||||   ||||||| |||||     | |||||||
Db   1049 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1108

Qy   1205 aaatcatccatctgatg 1221
          ||||  |    |||||
Db   1109 AAATTGTTTCCATGATG 1125


RESULT   4
US-09-073-674-4
; Sequence 4, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1140
US-09-073-674-4


  Query Match           35.0%;  Score 444.2;  DB 2;  Length 1140;
  Best Local Similarity   62.8%;  Pred. No. 1.5e-134;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    ||||||||| ||| | || |||    |    || || ||   |||||
Db     29 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 88

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          ||||||||||  |  || ||||| || ||||||||||    |  |||| ||||||   ||
Db     89 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 148

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    149 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 208

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |  || |  |||||||| |||||||||||||||| ||||||| |||||| ||||| |
Db    209 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 268

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||||| |||||||||||||||| ||| ||||| |  | |   |
Db    269 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 328

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
          |      |||    |||    |  || ||||||||||||||  | ||| ||||||||||  |
Db    329 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 388

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||||  ||     ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    389 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 448

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db    449 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 508

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    509 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 568

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||     | |||| |  || |  | ||      |  |  |||||     | |||||
Db    569 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 628

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||       ||  .  |   |||     |   || |  |   |  | ||   ||
Db    629 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 688

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    689 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 748

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  ||||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    749 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 808

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
```

```
Db     809 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 868

Qy     965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db     869 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 928

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db     929 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 988

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db     989 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1048

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db    1049 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1108

Qy    1205 aaatcatccatctgatg 1221
Db    1109 AAATTGTTTCCATGATG 1125
```

RESULT 5
US-08-227-108-6
; Sequence 6, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1146 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
US-08-227-108-6

```
  Query Match            35.0%;  Score 444.2;  DB 1;  Length 1146;
  Best Local Similarity  62.8%;  Pred. No. 1.5e-134;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy     125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db      26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy     185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
Db      86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy     245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db     146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy     305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db     206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy     365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db     266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGCCAGGAGGAATC 325

Qy     425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db     326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy     485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
Db     386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy     545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db     446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy     605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
Db     506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy     665 aaagccccgggaccaaattttttgttgctgccagatatgatgatcaagggattgtttggca 724
Db     566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
Db     626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db     686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy     845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db     746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db     806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy     965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db     866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db     926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db     986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db    1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy    1205 aaatcatccatctgatg 1221
```

```
Db   1106 AAATTGTTTCCATGATG 1122


RESULT   6
US-09-073-674-6
; Sequence 6, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1146 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-09-073-674-6


  Query Match             35.0%;  Score 444.2;  DB 2;  Length 1146;
  Best Local Similarity   62.8%;  Pred. No. 1.5e-134;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    |||||||||| ||| | || |||    |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          |||||||||| |  || ||||| || |||||||||    |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
             |  |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |   || |  |||||||| |||||||||||||||| ||||||| |||||| ||||| |
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||| ||||||||||||||| ||| ||||| |  | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
            |     |||    |||   |  || |||||||||||||| | ||| ||||||||||  |

Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||||| ||    ||  ||| || ||| |||||||||  |  | ||  |  |  |
Db    386 ATGACCTTGCCGCCACCATTGACTTCNTCTTGAAGAAAACGGGACAGGACAAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          | || ||| |||| ||||||||||||||| || || || || ||||||||||  ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||  |   |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||     | |||| |  || |  | ||      |  | |||||    | ||||| |
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784
          | |||  |||  ·     ||  |  |||      |   || |  |  | | ||  ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||   | ||||    | || || ||  | ||||  |  |  |||||||  |||| |
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  |||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | || ||  | || ||||||  |||||| || || ||||  |  |||  || |||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| ||         | ||  ||    | |  |||||   |  |||     ||||
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| ||| |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          | | || ||  |  |  |||| ||  || |  |||| |||| |||| | ||| ||||
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
          ||   |    |||| |||| || |||||||   |||||| ||||     | |||||||
Db    1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1205 aaatcatccatctgatg 1221
          ||||  |     |||||
Db    1106 AAATTGTTTCCATGATG 1122


RESULT   7
US-08-227-108-1
; Sequence 1, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1528 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-227-108-1


  Query Match             35.0%;  Score 444.2;  DB 1;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 1.8e-134;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy     125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
           |||| ||||    |||||||||| ||| | || |||    |    || || ||   |||||
Db      26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy     185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
           |||||||||| |  || ||||| || |||||||||    |  |||| ||||||   ||
Db      86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy     245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
           |  |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db     146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy     305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |   || |  |||||||| |||||||||||||||||| ||||||| |||||| ||||| |
Db     206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy     365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
           | || || |  |||||||| |||||||||||||||| ||| ||||| |  | |   |
Db     266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy     425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
           |     |||    |||   |  || ||||||||||||||| | ||| ||||||||||  |
Db     326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy     485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
           |||||||||| ||    ||  |||| || ||| ||||||||||| ||||| ||| |  | |
Db     386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy     545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
           | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db     446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy     605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
           ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db     506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy     665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
           ||| ||     | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db     566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
           | |||  |||       ||   |  |||    |  || | |  |  | | ||   ||
Db     626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
           |  ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db     686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy     845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
           |  ||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db     746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
           | || ||  | || ||||||  |||||| || || |||||  |  |||  || |||||||
Db     806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy     965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
           |||| ||         | ||  ||    | |   |||||   |  |||      ||||
Db     866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db     926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db     986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           ||   |    |||| |||| || |||||||  |||||| |||||     | |||||||
Db    1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy    1205 aaatcatccatctgatg 1221
           ||||  |     |||||
Db    1106 AAATTGTTTCCATGATG 1122


RESULT   8
US-09-073-674-1
; Sequence 1, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1528 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-09-073-674-1


  Query Match             35.0%;  Score 444.2;  DB 2;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 1.8e-134;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    |||||||| ||| | || |||     |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          |||||||||| |  || ||||| || |||||||||     |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
             |  | |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||| |||||||||||||||| |||||| |||||| ||||| |
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||| ||||||||||||||| ||| ||||| |  | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
           |     |||    |||   |  || |||||||||||||| | ||| ||||||||||  |
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcatttttccaccatgccagagc 604
          | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||  |   ||||||| |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||     | |||| |  || |  | ||     |  | |||||    | ||||| |
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||        ||   |  |||      |   || |  |   |  | ||   ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  |||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | || ||  | || ||||||  |||||| || || |||||  |  |||  || |||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          ||||  ||        | ||  ||    | |   |||||   |  |||      ||||
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
             ||   |    |||| |||| || |||||||   |||||||  |||||     | |||||||
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1205 aaatcatccatctgatg 1221
          ||||  |    |||||
Db   1106 AAATTGTTTCCATGATG 1122


RESULT   9
US-08-751-782-2
; Sequence 2, Application US/08751782
; Patent No. 5821352
; GENERAL INFORMATION:
;   APPLICANT:  Heintz, Nathaniel
;   APPLICANT:  Gubbay, Johnathan
;   APPLICANT:  Skinner, Michael
;   TITLE OF INVENTION:  A cDNA Library Prepared during
;   TITLE OF INVENTION:  Regression of Rat Prostate and Methods of Use Thereof
;   NUMBER OF SEQUENCES:  15
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  David A. Jackson, Esq.
;     STREET:  411 Hackensack Ave, Continental Plaza, 4th
;     STREET:  Floor
;     CITY:  Hackensack
;     STATE:  New Jersey
;     COUNTRY:  USA
;     ZIP:  07601
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/751,782
;     FILING DATE:  18-NOV-1995
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Jackson Esq., David A.
;     REGISTRATION NUMBER:  26,742
;     REFERENCE/DOCKET NUMBER:  600-1-190
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  201-487-5800
;     TELEFAX:  201-343-1684
; INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  178 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   HYPOTHETICAL:  NO
;   IMMEDIATE SOURCE:
;     CLONE:  10.2
US-08-751-782-2


  Query Match              3.8%;  Score 48.2;  DB 1;  Length 178;
  Best Local Similarity   55.8%;  Pred. No. 3.7e-06;
  Matches   92;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    416 gaaaacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgaga 475
          | |||||   |   ||   |  |||     | || || |||| ||| |||| | ||||
Db      8 GGAAACATGTGCGCCTAGACCCAGATTCTAAAGAATTTTGGGATTTTAGTTTTAATGAAC 67

Qy    476 tggctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaa 535
                  |  ||||| || ||  | ||  | || ||| || |  ||||  | ||   |
Db     68 AAATAGAATACGACCTCCCAGCCATCATTTATTTCATTCTGAATGAAACAAGACAAACAC 127
```

```
Qy    536 agatctattatgtcggctattcacagggcaccaccatgggcttta 580
          | || || ||| | ||| |||| ||||||  |    || |  | |
Db    128 AAATTTACTATATTGGCCATTCCCAGGGCGTCTATCTGCGTATCA 172


RESULT  10
US-08-925-171-2
; Sequence 2, Application US/08925171
; Patent No. 5928871
;  GENERAL INFORMATION:
;    APPLICANT:  Heintz, Nathaniel
;    APPLICANT:  Gubbay, Johnathan
;    APPLICANT:  Skinner, Michael
;    TITLE OF INVENTION:  A cDNA Library Prepared during
;    TITLE OF INVENTION:  Regression of Rat Prostate and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/925,171
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-190 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  178 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  10.2
US-08-925-171-2


  Query Match             3.8%;  Score 48.2;  DB 2;  Length 178;
  Best Local Similarity   55.8%;  Pred. No. 3.7e-06;
  Matches   92;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    416 gaaaacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgaga 475
          | |||||   |   ||   |  |||     | || || |||| ||| |||| | ||||
Db      8 GGAAACATGTGCGCCTAGACCCAGATTCTAAAGAATTTTGGGATTTTAGTTTTAATGAAC 67

Qy    476 tggctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaa 535
          |  |||| || ||  | ||  | || ||| || |  ||||  | ||   |
Db     68 AAATAGAATACGACCTCCCAGCCATCATTTATTTCATTCTGAATGAAACAAGACAAACAC 127

Qy    536 agatctattatgtcggctattcacagggcaccaccatgggcttta 580
          | || || ||| | ||| |||| ||||||  |    || |  | |
Db    128 AAATTTACTATATTGGCCATTCCCAGGGCGTCTATCTGCGTATCA 172


RESULT  11
US-09-356-952-12
; Sequence 12, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
;  APPLICANT: Boriack-Sjodin, Ann
;  APPLICANT: Margarit, S. M.
;  APPLICANT: Bor-Sogi, Dafna
;  APPLICANT: Cole, Philip
;  APPLICANT: Kuriyan, John
;  TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: 600-1-228N
;  CURRENT APPLICATION NUMBER: US/09/356,952
;  CURRENT FILING DATE: 1999-07-19
;  EARLIER APPLICATION NUMBER: 60/093,631
;  EARLIER FILING DATE: 1998-07-21
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 43676
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-356-952-12


  Query Match             3.7%;  Score 46.4;  DB 3;  Length 43676;
  Best Local Similarity   46.0%;  Pred. No. 0.00035;
  Matches  232;  Conservative    0;  Mismatches  266;  Indels    6;  Gaps    2;

Qy     94 aattcagtacatatgccaactaaagctgtggacccagaagcattcatgaatattagtgaa 153
          ||| || |||| |  | ||    | | |  |||     |  | |   | || ||| | |
Db  14682 aatccattacagaatctaatcttatcggaagacttgacattagttgcggatcttaattat 14741

Qy    154 atcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggt-- 211
            | | ||  |      ||     |||||||  |  | | |  || ||||||||||
Db  14742 tattttaaccagtacaatattcaaatagaggaatttagattggagaccgaagatgggttt 14801

Qy    212 -atatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttcc 270
           |||   ||      ||  ||| ||   |     |  | |  | |    | |
Db  14802 gttatagatttgtggcacttgataccaaaatatagaacgacagattctgacaagaagaag 14861

Qy    271 aggcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaac 330
          |||||     | ||  |||  |||||||| |   | |   | |    |    |||||
Db  14862 aggccacccattttgatgctacatggccttttgcaaagcagtggttcgttcgcatccaa- 14920

Qy    331 ctgcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatgggg 390
              |  ||    ||||  |   |||||     |  | ||||  ||| | ||| | |||
Db  14921 --tggtagaaaatctctggcatattttctgtatcaatccggttacgacatatggttaggg 14978

Qy    391 aacagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgag 450
          || | |||  |        || |   |    || |  |             | || |
Db  14979 aataacagatgcgggtttaggccggaatggaacgaagcgaaagtaccgacactagcttcc 15038

Qy    451 ttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaacttt 510
            ||||  |       |   ||||||| || ||  || ||  |    ||||  |
Db  15039 aggtgggactgggaccttcgcgagatggttaagtacgatctgacccttttgattgatacc 15098

Qy    511 attttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccacc 570
            | ||    || |||     |||||| |   ||   ||    |||| |||||||| ||
Db  15099 gtgttagctaagacgcagtttgaaaagcttactttgatctcgcattctcagggcactaca 15158

Qy    571 atgggctttattgcattttccacc 594
            ||| |||||  |  ||   || |
Db  15159 caggggtttatgggcttggtcaac 15182


RESULT  12
US-08-852-806-1
; Sequence 1, Application US/08852806
; Patent No. 5874245
;  GENERAL INFORMATION:
;    APPLICANT:  Shoji Fukusumi
;    APPLICANT:  Shuji Hinuma
;    TITLE OF INVENTION:  NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR
;    NUMBER OF SEQUENCES:  14
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/852,806
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/017,915
;      FILING DATE:  16 MAY 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Han, William T
;      REGISTRATION NUMBER:  34,344
;      REFERENCE/DOCKET NUMBER:  TAK50002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5219
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3271 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-852-806-1


  Query Match             2.9%;  Score 36.6;  DB 2;  Length 3271;
  Best Local Similarity   50.9%;  Pred. No. 0.12;
  Matches   87;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     969 gagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagagtcag 1028
           || ||  ||  | ||||| |  |   ||| ||||| |   || ||           |||
Db     678 GATTGTCACAGATAATCTAGTGATCAGCATTCAGCGAGAGCCCGTCTCAGCTGTGTCCAG 737

Qy    1029 agatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatccaga 1088
           || || ||| |||| |          |    ||     |||||||| |     |  ||||
Db     738 TGACATCACGTTCCCCATGCGGGGCCGCCGGGGCATGAAGGACTGGGTGCGGCACTCAGA 797

Qy    1089 agacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaa 1139
           |||       | |||| |   ||||||  || ||||  || ||    |||
Db     798 GGACCGCCTCTTCCTGCCCAAGGAGGTGCTCAGCCTCTCCTCCCCAGGGAA 848


RESULT  13
US-09-163-669-1
; Sequence 1, Application US/09163669
; Patent No. 6111076
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUSUMI, SHOJI
;    APPLICANT:  HINUMA, SHUJI
;    APPLICANT:  FUJII, RYO
;    TITLE OF INVENTION:  NOVEL HUMAN G-PROTEIN COUPLED
;    TITLE OF INVENTION:  RECEPTOR (HIBCD07)
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ratner & Prestia
;      STREET:  P.O. Box 980
;      CITY:  Valley Forge
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/163,669
;      FILING DATE:  30-SEP-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/852,806
;      FILING DATE:  07-MAY-1997
;      APPLICATION NUMBER:  60/017,915
;      FILING DATE:  16-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Prestia, Paul F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  TAK-50002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0700
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3271 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-163-669-1


  Query Match             2.9%;  Score 36.6;  DB 3;  Length 3271;
  Best Local Similarity   50.9%;  Pred. No. 0.12;
  Matches   87;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     969 gagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagagtcag 1028
           || ||  ||  | ||||| |  |   ||| ||||| |   || ||           |||
Db     678 GATTGTCACAGATAATCTAGTGATCAGCATTCAGCGAGAGCCCGTCTCAGCTGTGTCCAG 737

Qy    1029 agatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatccaga 1088
           || || ||| |||| |          |    ||     |||||||| |     |  ||||
Db     738 TGACATCACGTTCCCCATGCGGGGCCGCCGGGGCATGAAGGACTGGGTGCGGCACTCAGA 797

Qy    1089 agacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaa 1139
           |||       | |||| |   ||||||  || ||||  || ||    |||
Db     798 GGACCGCCTCTTCCTGCCCAAGGAGGTGCTCAGCCTCTCCTCCCCAGGGAA 848


RESULT  14
US-08-276-852-154
; Sequence 154, Application US/08276852
; Patent No. 5652138
;  GENERAL INFORMATION:
;    APPLICANT:  Burton, Dennis R
;    APPLICANT:  Barbas, Carlos F
;    APPLICANT:  Lerner, Richard A
;    TITLE OF INVENTION:  HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
;    TITLE OF INVENTION:  TO HUMAN IMMUNODEFICIENCY VIRUS
;    NUMBER OF SEQUENCES:  170
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10666 No. 5652138th Torrey Pines Road, Suite 220,
;      STREET:  Mail Drop TPC8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,852
;      FILING DATE:  18-JUL-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/178,302
;      FILING DATE:  30-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/954,148
;      FILING DATE:  30-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  SCR1452P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  154:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  15..452
US-08-276-852-154


  Query Match             2.9%;  Score 36.6;  DB 1;  Length 3282;
  Best Local Similarity   51.5%;  Pred. No. 0.12;
  Matches   84;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy    435 catagaccaagatgagttctgggctttcagttatgatgagatggctaggtttgaccttcc 494
          || || | |    |||||  ||    |||| | ||  || | ||||   | | |   | |
Db    302 CACAGCCTACATGGAGTTGAGGAGCCTCAGGTCTGCAGACACGGCTGTTTATTATTGTGC 361

Qy    495 tgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctattatgtcggcta 554
            ||||      | || || | |  |  |   |||||| ||    |||  | || ||
Db    362 GAGAGTGGGGCCATATAGTTGGGATGATTCTCCCCAGGACAATTATTATATGGACGTCTG 421

Qy    555 ttcacagggcaccaccatgggctttattgcattttccaccatg 597
              | || |||||  |   | | |   ||  ||||||||
Db    422 GGGCAAAGGAACCACGGTCATCGTGAGCTCAGCTTCCACCAAG 464


RESULT  15
US-08-276-852-169/c
; Sequence 169, Application US/08276852
; Patent No. 5652138
;  GENERAL INFORMATION:
;    APPLICANT:  Burton, Dennis R
;    APPLICANT:  Barbas, Carlos F
;    APPLICANT:  Lerner, Richard A
;    TITLE OF INVENTION:  HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
;    TITLE OF INVENTION:  TO HUMAN IMMUNODEFICIENCY VIRUS
;    NUMBER OF SEQUENCES:  170
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10666 No. 5652138th Torrey Pines Road, Suite 220,
;      STREET:  Mail Drop TPC8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,852
;      FILING DATE:  18-JUL-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/178,302
;      FILING DATE:  30-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/954,148
;      FILING DATE:  30-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  SCR1452P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  169:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-276-852-169


  Query Match             2.9%;  Score 36.6;  DB 1;  Length 3282;
  Best Local Similarity   51.5%;  Pred. No. 0.12;
  Matches   84;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy    435 catagaccaagatgagttctgggctttcagttatgatgagatggctaggtttgaccttcc 494
          || || | |    |||||  ||    |||| | ||  || | ||||   | | |   | |
Db   2981 CACAGCCTACATGGAGTTGAGGAGCCTCAGGTCTGCAGACACGGCTGTTTATTATTGTGC 2922

Qy    495 tgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctattatgtcggcta 554
            ||||      | || || | |  |  |   |||||| ||    |||  | || ||
Db   2921 GAGAGTGGGGCCATATAGTTGGGATGATTCTCCCCAGGACAATTATTATATGGACGTCTG 2862

Qy    555 ttcacagggcaccaccatgggctttattgcattttccaccatg 597
              | || |||||  |   | | |   ||  ||||||||
Db   2861 GGGCAAAGGAACCACGGTCATCGTGAGCTCAGCTTCCACCAAG 2819


Search completed: January 30, 2002, 22:11:14
Job time: 6251 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 30, 2002, 20:23:58 ; Search time 1536.27 Seconds
(without alignments)
8876.300 Million cell updates/sec

Title: US-09-578-063-46
Perfect score: 1269
Sequence: 1 atgttggaaaccttgtcaag.....gacggtgtgaggccgtattg 1269

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	618.2	48.7	2927	12	AK019504	AK019504 Mus muscu
2	505	39.8	954	11	BF135102	BF135102 601780142
3	497.6	39.2	1079	11	BF141567	BF141567 601788654
4	471.8	37.2	669	11	BF143546	BF143546 601790896
5	461.8	36.4	1350	12	AK010093	AK010093 Mus muscu
6	460.2	36.3	1347	12	AK010116	AK010116 Mus muscu
7	460.2	36.3	1349	12	AK009300	AK009300 Mus muscu
8	460.2	36.3	1350	12	AK010139	AK010139 Mus muscu
9	458.6	36.1	1340	12	AK010124	AK010124 Mus muscu
10	458.6	36.1	1343	12	AK009479	AK009479 Mus muscu
11	458.6	36.1	1345	12	AK009413	AK009413 Mus muscu
12	458.6	36.1	1345	12	AK009473	AK009473 Mus muscu
13	458.6	36.1	1346	12	AK009359	AK009359 Mus muscu
14	458.6	36.1	1346	12	AK009523	AK009523 Mus muscu
15	458.6	36.1	1346	12	AK009729	AK009729 Mus muscu
16	458.6	36.1	1348	12	AK010058	AK010058 Mus muscu
17	458.6	36.1	1349	12	AK009474	AK009474 Mus muscu
18	458.6	36.1	1349	12	AK009546	AK009546 Mus muscu
19	458.6	36.1	1349	12	AK009571	AK009571 Mus muscu
20	458.6	36.1	1349	12	AK009573	AK009573 Mus muscu
21	458.6	36.1	1349	12	AK010148	AK010148 Mus muscu
22	458.6	36.1	1350	12	AK009459	AK009459 Mus muscu
23	458.6	36.1	1350	12	AK009525	AK009525 Mus muscu
24	458.6	36.1	1350	12	AK009773	AK009773 Mus muscu
25	458.6	36.1	1350	12	AK010061	AK010061 Mus muscu
26	458.6	36.1	1350	12	AK010125	AK010125 Mus muscu
27	458.6	36.1	1350	12	AK010203	AK010203 Mus muscu
28	458.6	36.1	1351	12	AK009428	AK009428 Mus muscu
29	458.6	36.1	1352	12	AK009431	AK009431 Mus muscu
30	458.6	36.1	1356	12	AK010019	AK010019 Mus muscu
31	458.6	36.1	1356	12	AK010035	AK010035 Mus muscu
32	457.6	36.1	1346	12	AK009437	AK009437 Mus muscu
33	457.6	36.1	1347	12	AK010103	AK010103 Mus muscu
34	457.6	36.1	1349	12	AK010026	AK010026 Mus muscu
35	457	36.0	1346	12	AK009560	AK009560 Mus muscu
36	457	36.0	1347	12	AK010236	AK010236 Mus muscu
37	457	36.0	1348	12	AK010231	AK010231 Mus muscu
38	457	36.0	1349	12	AK009875	AK009875 Mus muscu
39	457	36.0	1350	12	AK009537	AK009537 Mus muscu
40	456	35.9	1349	12	AK010106	AK010106 Mus muscu
41	455.4	35.9	1345	12	AK009544	AK009544 Mus muscu
42	455.4	35.9	1349	12	AK009990	AK009990 Mus muscu
43	453.8	35.8	1346	12	AK010173	AK010173 Mus muscu
44	449.8	35.4	1407	12	AK009923	AK009923 Mus muscu
45	448.2	35.3	1350	12	AK010123	AK010123 Mus muscu

ALIGNMENTS

RESULT 1
AK019504
LOCUS AK019504 2927 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus 0 day neonate skin cDNA, RIKEN full-length enriched library, clone:4632427C23, full insert sequence.
ACCESSION AK019504
VERSION AK019504.1 GI:12859754
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 0 day neonate skin cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:4632427C23.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2927)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 2927)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (bases 1 to 2927)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,

```
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 2927)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2927)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 10.0 and subtraction to
            Rot = 100.0. Second strand cDNA was prepared with the primer
            adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. cDNA of size comprised between 3 and 7 kb was
            selected before cloning. Vector: a modified pBluescript KS(+) after
            bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES                Location/Qualifiers
     source             1. .2927
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /db_xref="MGD:MGI:1912763"
                        /db_xref="MGD:MGI:1926003"
                        /clone="4632427C23"
                        /tissue_type="skin"
                        /clone_lib="RIKEN full-length enriched mouse cDNA library"
                        /dev_stage="0 day neonate"
     CDS                153. .662
                        /note="putative"
                        /codon_start=1
                        /protein_id="BAB31766.1"
                        /db_xref="GI:12859755"
                        /translation="MSEILSRVWTVSHRVEIWLLILVAYLLQRNVNSGHLPTKAADPE
                        AFMNVSEIIKHKGYPSEEYEVATEDGYILSVNRIPRGQTRLKKEGSRPVVLLQHGLLG
                        DASNWISNLPNNSLGFILADAGFDVWMGNSRGNTWSRKHKTLSIDQDEFWAFRYKNTL
                        RVEVDTDMP"
     polyA_signal       2908. .2913
                        /note="putative"
     polyA_site         2927
                        /note="putative"

BASE COUNT      837 a    581 c    664 g    845 t
ORIGIN

  Query Match             48.7%;  Score 618.2;  DB 12;  Length 2927;
  Best Local Similarity   85.3%;  Pred. No. 2.1e-169;
  Matches  702;  Conservative    0;  Mismatches  118;  Indels    3;  Gaps    1;

Qy     447 tgagttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaa 506
           ||  || |   ||  |||||||||| ||||||||||||||||||||| || ||||||||
Db    1376 TGTCTTTTCCCTTTGTAGTTATGATGAAATGGCTAGGTTTGACCTTCCAGCTGTGATAAA 1435

Qy     507 ctttattttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcac 566
           |||||| || |||||||||||||| ||||| |||||||||||||| ||||||||||
Db    1436 CTTTATCTTACAGAAAACGGGCCAGAAAAAGGTCTATTATGTCGGCTACTCACAGGGCAC 1495

Qy     567 caccatgggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgta 626
           |||||||||||||||||||||||||| |||||||||| |||| ||||||||||||||
Db    1496 CACCATGGGCTTTATTGCATTTTCCACAATGCCAGAGCTAGCTCATAAAATCAAAATGTA 1555

Qy     627 ttttgctttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaattttt 686
           |||||| ||||| || |||||||||||||  |||||| ||| || || ||||||||| |
Db    1556 TTTTGCCTTAGCTCCTATAGCCACTGTTAAATATGCAAGAAGTCCTGGTACCAAATTTCT 1615

Qy     687 gttgctgccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagac 746
           | |||||||||||||||||||||||| ||| ||||||| | |||| ||| | || |||||
Db    1616 GCTGCTGCCAGATATGATGATCAAGGTATTATTTGGCAGACAAGAGTTTTTATACCAGAC 1675

Qy     747 cagatttctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttg 806
           |||||| ||||||| ||| |||||||||||||||||||| |||||||||| || || ||
Db    1676 TAGATTTTTCAGACAGCTTTTTATTTACCTTTGTGGCCAGATGATTCTTGACCAAATCTG 1735

Qy     807 tagtaatatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagc 866
           || || ||||| ||||| ||||| || || ||||| ||||||||||||||||||| ||||
Db    1736 CAGCAACATCATCTTACTCCTGGGAGGTTTTAACACAAACAATATGAACATGAGCAGAGC 1795

Qy     867 aagtgtatatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggag 926
           || ||| ||||  ||||| || | |||||||| |||||||| || ||||| ||||||||
Db    1796 AAATGTGTATGTCGCCCATACACCTGCTGGAACGTCTGTGCAGAACATTCTCCACTGGAG 1855

Qy     927 ccaggcagtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatct 986
           |||||| |||||||||| ||||| || || |||||||||| ||||||||||||||||||
Db    1856 CCAGGCCGTGAATTCTGGGGAACTTCGTGCCTTTGACTGGGGAAGTGAGACCAAAAATCA 1915

Qy     987 ggaaaaatgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctac 1046
           ||| ||||| |||||||||||||| ||||||||| |||  ||||||||| |||||| ||
Db    1916 GGAGAAATGTAATCAGCCAACTCCTATAAGGTACAAAGTTCGAGATATGATGGTCCCAAC 1975

Qy    1047 agcaatgtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgct 1106
           |||||||||||| |||||||| |||||||||||||||||||| || |||||||   | ||
Db    1976 AGCAATGTGGACTGGAGGTCAAGACTGGCTTTCAAATCCAGATGATGTGAAAACATTACT 2035

Qy    1107 ctctgaggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggattt 1166
           ||||| || |||||||||||||||| ||||| |||||||| |||||||| ||||||||
Db    2036 TTCTGAAGTAACCAACCTCATCTACCACAAGAACATTCCTGAGTGGGCTCATGTGGATTT 2095

Qy    1167 catctggggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagca 1226
           |||||||||  ||||||| ||||| ||| | |||||||||||| || ||||||||| |||
Db    2096 CATCTGGGGGCTGGATGCCCCTCAGCGTGTTTACAATGAAATTATACATCTGATGAAGC- 2154

Qy    1227 ggaggagaccaaccttt cccagggacggtgtgaggccgtattg 1269
           |||||  |||||||| ||||||||    ||   || |  ||||
Db    2155 --AGGAGCCCAACCTTCCCCAGGGAACCTGCAGGGTCAAATTG 2195


RESULT   2
BF135102
LOCUS       BF135102      954 bp    mRNA            EST       24-OCT-2000
DEFINITION  601780142F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4008131 5',
            mRNA sequence.
ACCESSION   BF135102
VERSION     BF135102.1  GI:10974142
KEYWORDS    EST.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 954)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9242  row: c  column: 12
            High quality sequence stop: 695.
FEATURES             Location/Qualifiers
     source          1. .954
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4008131"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      280 a    222 c    222 g    230 t
ORIGIN


  Query Match           39.8%;  Score 505;  DB 11;  Length 954;
  Best Local Similarity   85.3%;  Pred. No. 1.9e-136;
  Matches  598;  Conservative    0;  Mismatches  100;  Indels    3;  Gaps    3;

Qy    493 cctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctattatgtcggc 552
          || || |||||||||||| || || ||||||||||||||| ||||| ||||||||||||||
Db      1 CCAGCTGTGATAAACTTGATCTTACAGAAAACGGGCCAGAAAAAGGTCTATTATGTCGGC 60

Qy    553 tattcacagggcaccaccatgggctttattgcattttccaccatgccagagctggctcag 612
          || |||||||||||||||||||||||||||||||||||||||| |||||||||||| |||||
Db     61 TACTCACAGGGCACCACCATGGGCTTTATTGCATTTTCCACAATGCCAGAGCTAGCTCAT 120

Qy    613 aaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaaaaagcccc 672
          |||||||||||||||||| ||||| || |||||||||||||  |||||| ||| ||
Db    121 AAAATCAAAATGTATTTTGCCTTAGCTCCTATAGCCACTGTTAAATATGCAAGAAGTCCT 180

Qy    673 gggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggcaaaaaagaa 732
          || ||||||||| || ||||||||||||||||||||||| ||| ||||||| | ||||
Db    181 GGTACCAAATTTCTGCTGCTGCCAGATATGATGATCAAGGTATTATTTGGCAGACAAGAG 240

Qy    733 tttctgtatcagaccagatttctcagacaacttgttatttacctttgtggccaggtgatt 792
          ||| | || |||||| |||||| |||| || ||| |||||||||| |||||||||| |||||
Db    241 TTTTTATACCAGACTAGATTTTTCAGGCAGCTTTTTATTTACC-TTGTGGCCAGATGATT 299

Qy    793 cttgatcagatttgtagtaatatcatgttacttctgggtggattcaacaccaacaatatg 852
          || || || || || || ||||| ||||| ||||| || || ||||| |||||||||
Db    300 CTCGACCAAATCTGCAGCAACATCATCTTACTCCTGGGAGGTTTTAACACAAACAATATG 359

Qy    853 aacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctgtgcaaaat 912
          ||||||||| |||||| ||| |||| ||||| || | |||||||||| |||||||| ||
Db    360 AACATGAGCAGAGCAAATGTGTATGTCGCCCATACACCTGCTGGAACGTCTGTGCAGAAC 419

Qy    913 attctacactggagccaggcagtgaattctggtgaactccgggcatttgactgggggagt 972
          ||||| |||||||||||||| |||||||||| ||||| || || ||||||||||| |||
Db    420 ATTCTCCACTGGAGCCAGGCCGTGAATTCTGGGGAACTTCGTGCCTTTGACTGGGGAAGT 479

Qy    973 gagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagagtcagagat 1032
          |||||||||||| ||| ||||| |||||||||||||| ||||||||| |||  |||||
Db    480 GAGACCAAAAATCAGGAGAAATGTAATCAGCCAACTCCTATAAGGTACAAAGTTCGAGAT 539

Qy   1033 atgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatccagaagac 1092
          |||| |||||| |||||||||||||| |||||||| |||||| |||||||||||| ||
Db    540 ATGATGGTCCCAACAGCAATGTGGACTGGAGGTCAAGACTGG-TTTCAAATCCAGATGAT 598

Qy   1093 gtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattcctgaatgg 1152
          ||||||    | || ||||| || ||||| |||||||||| ||||| ||||||||   |
Db    599 GTGAAACAATTACTTTCTGAAGTAACCAAACTCATCTACCACAAGAACATTCCTGAGGTG 658

Qy   1153 gctcacgtggatttcatctggggtttggatgctcctcaccg 1193
          | ||| ||||| |||||||||| ||||||| ||||| ||
Db    659 GGTCATGTGGA-TTCATCTGGGGGGTGGATGCCCCTCAGCG 698


RESULT   3
BF141567
LOCUS       BF141567     1079 bp    mRNA            EST       24-OCT-2000
DEFINITION  601788654F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4016413 5',
            mRNA sequence.
ACCESSION   BF141567
VERSION     BF141567.1  GI:10980607
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1079)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9263  row: l  column: 14
            High quality sequence stop: 663.
FEATURES             Location/Qualifiers
     source          1. .1079
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4016413"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      312 a    267 c    263 g    237 t
ORIGIN


  Query Match           39.2%;  Score 497.6;  DB 11;  Length 1079;
  Best Local Similarity   84.1%;  Pred. No. 2.8e-134;
  Matches  621;  Conservative    0;  Mismatches  109;  Indels    8;  Gaps    5;

Qy    256 aagaagacaggttccaggcctgtggtgttactgcagcatggcctagttggaggtgctagc 315
          |||||| ||| |||||||| |||||||||||||| ||||| ||  |||| | |||||||
Db     20 AAGAAGGAAGGATCCAGGCCAGTGGTGTTACTGCAACATGGTCTTCTTGGGGATGCTAGC 79

Qy    316 aactggatttccaacctgcccaacaatagcctgggcttcattctggcagatgctggtttt 375
          ||||||||||||||||||||||||||| |||||||| || ||||||||||||| ||||||
Db     80 AACTGGATTTCCAACCTGCCCAACAACAGCCTGGGTTTTATTCTGGCAGATGCAGGTTTT 139
```

```
Qy    376 gacgtgtggatggggaacagcaggggaaacgcc-tggtctcgaaaacacaagacactctc 434
          || |||||||||| |||||||| |||||| || |||||||| || |||||||| |||||
Db    140 GATGTGTGGATGGGAAACAGCAGAGGAAACACCGTGGTCTCGGAAGCACAAGACCCTCTC 199

Qy    435 catagaccaagatgagttctgggctttcagttatgatgagatggctaggtttgaccttcc 494
          |||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    200 CATAGATCAAGATGAGTTCTGGGCTTTCAGTTATGATGAAATGGCTAGGTTTGACCTTCC 259

Qy    495 tgcagtgataaactttattttgcagaaaacgggccaggaaaagatctattatgtcggcta 554
          || |||||||||||||| || |||||||||||||| ||||| ||||||||||||||||
Db    260 AGCTGTGATAAACTTTATCTTACAGAAAACGGGCCAGAAAAAGGTCTATTATGTCGGCTA 319

Qy    555 ttcacagggcaccaccatgggctttattgcattttccaccatgccagagctggctcagaa 614
          |||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||| ||
Db    320 CTCACAGGGCACCACCATGGGCTTTATTGCATTTTCCACAATGCCAGAGCTAGCTCATAA 379

Qy    615 aatcaaaatgtattttgctttagcacccatagccactgttaaagcatgcaaaaagccccgg 674
          |||||||||||||||| || ||||| || |||||||||||||  |||||| ||| || ||
Db    380 AATCAAAATGTATTTGGCCTTAGCTCCTATAGCCACTGTTAAATATGCAAGAAGTCCTGG 439

Qy    675 gaccaaatttttgttgctgccagatatgatgatcaagggattgtttggcaaaaaagaatt 734
          ||||||||| || |||||||||||||||||||||||| ||| ||||||| | |||| ||
Db    440 TACCAAATTTCTGCTGCTGCCAGATATGATGATCAAGGTATTATTTGGCAGACAAGAGTT 499

Qy    735 tctgtatcagaccagatttctcagacaacttgttatttaccttttgtggccaggtgattct 794
          | | || ||||| ||||| |||| || ||| |||||||||||||||| ||| |||||||
Db    500 TTTATACCAGACTAGATTTTTCAGGCAGCTTTTTATTTACCTTTGTGGGCAGATGATTCT 559

Qy    795 tgatcagatttgtagtaatatcatgttacttctgggtggattcaacaccaacaatatgaa 854
          || || || || || || ||||| ||||| ||||| || || ||||| ||| |||||||
Db    560 CGACCAAATCTGCAGCAACATCATCTTACTCCTGGGAGGTTTTAACACAAACCATATGAA 619

Qy    855 catgagccgagca--agtgtatatgctgcccacactcttgctggaacatctgtg---caa 909
          ||||||| |||||  |     | ||  ||||| |  | ||||||||| |||  |   | |
Db    620 CATGAGCAGAGCAACATGTGCTCTGTCGCCCATAAACCTGCTGGAACGTCTTGGTGCCGA 679

Qy    910 aatattctacactggagccagg-cagtgaattctggtg-aactccgggcatttgactggg 967
          |  ||||| |||||||||||| | |||||||||| | | || || || |||||||| |
Db    680 ACCATTCTCCACTGGAGCCAGGCCGGTGAATTCTGGGGCACCTTCGTGCCTTTGACTGTG 739

Qy    968 ggagtgagaccaaaaatc 985
          ||   | ||| |  || |
Db    740 GGCAAGCGACAACCAAGC 757


RESULT   4
BF143546
LOCUS       BF143546      669 bp    mRNA            EST       24-OCT-2000
DEFINITION  601790896F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4021870 5',
            mRNA sequence.
ACCESSION   BF143546
VERSION     BF143546.1  GI:10982586
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9277  row: o  column: 23
            High quality sequence stop: 659.

FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4021870"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      195 a    140 c    165 g    169 t
ORIGIN


  Query Match             37.2%;  Score 471.8;  DB 11;  Length 669;
  Best Local Similarity   85.9%;  Pred. No. 8.7e-127;
  Matches  535;  Conservative    0;  Mismatches   87;  Indels    1;  Gaps    1;

Qy      1 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 60
          |||  |||| |||||||||  ||||| ||| || |||||| |||| || |||||||||
Db     48 ATGTCAGAAATCTTGTCAAGAGTGTGGACTGTTTCGCACAGAGTGGAGATATGGCTTCTG 107

Qy     61 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 120
          |||||||| || ||| |  |||| |||||||||| || | |||| |||| || ||||||
Db    108 ATTCTGGTAGCATATTTACTCCAAAGAAATGTGAACTCGGGACATTTGCCCACGAAAGCT 167

Qy    121 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 180
          | ||| ||||||||||||||||||| |||| |||||||||  ||||  | || |||||| ||
Db    168 GCGGATCCAGAAGCATTCATGAATGTTAGCGAAATCATCAAACACAAGGGTTATCCCAGT 227

Qy    181 gaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcga 240
          |||||  |||||||| |||||  |||||||||| |||||||||| |||||  || |||||
Db    228 GAGGAGTATGAAGTTGCAACAGAAGATGGGTACATCCTTTCTGTGAACAGAATCCCTCGG 287

Qy    241 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 300
          || |     |     ||||||  ||| |||||||| |||||||||||||| ||||| ||
Db    288 GGACAGACACGGTTAAAGAAGGAAGGATCCAGGCCAGTGGTGTTACTGCAACATGGTCTT 347

Qy    301 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 360
          |||| | |||||||||||||||||||||||||||||||||| |||||||| || ||||||
Db    348 CTTGGGGATGCTAGCAACTGGATTTCCAACCTGCCCAACAACAGCCTGGGTTTTATTCTG 407

Qy    361 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 420
          |||||||| || ||||| |||||||||| |||||||| |||||| |||||||||| ||
Db    408 GCAGATGCAGG-TTTGATGTGTGGATGGGAAACAGCAGAGGAAACACCTGGTCTCGGAAG 466

Qy    421 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 480
          |||||||| |||||||||| ||||||||||||||||||||||||||||||||| ||||||
Db    467 CACAAGACCCTCTCCATAGATCAAGATGAGTTCTGGGCTTTCAGTTATGATGAAATGGCT 526

Qy    481 aggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatc 540
          |||||||||||||| || |||||||||||||| || |||||||||||||| ||||| ||
Db    527 AGGTTTGACCTTCCAGCTGTGATAAACTTTATCTTACAGAAAACGGGCCAGAAAAAGGTC 586

Qy    541 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 600
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||| ||||||
Db    587 TATTATGTCGGCTACTCACAGGGCACCACCATGGGCTTTATTGCATTTTCCACAATGCCA 646

Qy    601 gagctggctcagaaaatcaaaat 623
          ||||| ||||| || ||| ||||
Db    647 GAGCTAGCTCATAACATCCAAAT 669


RESULT   5
AK010093
LOCUS       AK010093     1350 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310067K20, full insert sequence.
ACCESSION   AK010093
```

```
VERSION     AK010093.1  GI:12845297
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310067K20.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1350)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1350)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1350)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1350)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1350)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1350
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1898392"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310067K20"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53. .1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26697.1"
                     /db_xref="GI:12845298"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLVASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      396 a    290 c    286 g    378 t
ORIGIN
```

```
  Query Match          36.4%;  Score 461.8;  DB 12;  Length 1350;
  Best Local Similarity   63.8%;  Pred. No. 8.5e-124;
  Matches  700;  Conservative    0;  Mismatches  397;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||||  |||||||| ||||| | || ||     |    || |||||  ||||||
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          |||||||||| |  |||||||||||| || || ||    ||  | || ||||||   ||
Db    192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              | || |||| |  ||    || |||||||  |  ||||||||||  | ||||
Db    252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGGTTG 311

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||| | || ||||| ||||| |||||| |||||||| ||||
Db    312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          ||||||| | ||| |||||| |||||||||  | || ||  | ||||| || ||| |
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
              ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||| ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          |||| ||  | || |||||||||||| || || |||||||| ||||| ||||  || | |
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
Db    612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
```

```
Db    672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    972 GGGGGAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db   1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db   1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy   1205 aaatcatccatctgatg 1221
Db   1212 AGATAGTTACCATGATG 1228
```

```
RESULT   6
AK010116
LOCUS       AK010116     1347 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310068O05, full insert sequence.
ACCESSION   AK010116
VERSION     AK010116.1  GI:12845334
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310068O05.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1347)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1347)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1347)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1347)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1347)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES            Location/Qualifiers
     source         1. .1347
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /db_xref="MGD:MGI:1895801"
                    /db_xref="MGD:MGI:1914967"
                    /clone="2310068O05"
                    /sex="male"
                    /tissue_type="tongue"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="adult"
     CDS            51. .1238
                    /note="putative"
                    /codon_start=1
                    /protein_id="BAB26711.1"
                    /db_xref="GI:12845335"
                    /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                    SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                    FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                    GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                    FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFGV
                    YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYYDVSAMTVPI
                    AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
```

```
                    AED*
BASE COUNT      393 a    290 c    284 g    380 t
ORIGIN


  Query Match           36.3%;  Score 460.2;  DB 12;  Length 1347;
  Best Local Similarity 63.7%;  Pred. No. 2.5e-123;
  Matches  699;  Conservative    0;  Mismatches  398;  Indels    0;  Gaps    0;

Qy     125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
           |||| ||||||  ||||||| ||||| | || ||     |    || |||||  ||||||
Db     130 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 189

Qy     185 aatatgaagtcgcaactgaagatgggtatatccctttctgttaacaggattcctcgaggcc 244
           |||||||||| |  |||||||||| || || ||    ||  | || ||||||   ||
Db     190 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 249

Qy     245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
             |  || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db     250 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 309

Qy     305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |   || |  |||||||| | || |||||  |||||  ||||||| |||||||| ||||
Db     310 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 369

Qy     365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
           ||||||| | ||| |||||| |||||||||| | || ||  | ||||| || ||| |
Db     370 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 429

Qy     425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                 ||   ||||  || ||| ||||||||||||||| | ||||||  |||||||  |
Db     430 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 489

Qy     485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
           |||||||||  ||    ||| |||| ||| | ||||||||  || || || |||||  | |
Db     490 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 549

Qy     545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
           |||| ||  | || ||||||||||| || || |||||||| ||||| |||| || |  |
Db     550 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 609

Qy     605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
           ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
Db     610 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 669

Qy     665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
           |||| |||   |  ||  |||  |  | ||  |  |  |  ||||||   | ||||| |
Db     670 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 729

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
           | |||   ||  ||   || | |   ||       |  |||| ||   |  | ||   |
Db     730 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 789

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
           || | | || ||||   | || || ||      | ||  |  |  |||||||  |||  |
Db     790 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 849

Qy     845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
           | ||  | ||  |||| ||     |||| |||   |  || | ||  || ||||||||||
Db     850 AAAACTTAAATGTGAGTCGCTTTGGTGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 909

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
             ||| |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db     910 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 969

Qy     965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
           |||| |||       | ||  ||  | | | |||||||  ||| |||      ||    |
Db     970 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1029

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           |   ||  ||||| || || |  ||| ||||||  || || || |||   ||  |  |||
Db    1030 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1089

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||| | ||||
Db    1090 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1149

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
             |   |   |||| |||| ||||||||||   |||||||  |||||     | |||||||
Db    1150 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1209

Qy    1205 aaatcatccatctgatg 1221
           | ||  |     |||||
Db    1210 AGATAGTTACCATGATG 1226


RESULT   7
AK009300
LOCUS       AK009300     1349 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310011G18, full insert sequence.
ACCESSION   AK009300
VERSION     AK009300.1  GI:12844011
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310011G18.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1349)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1349)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1349)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1349)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1349)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
```

```
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1349
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1895751"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310011G18"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              53. .1240
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26201.1"
                      /db_xref="GI:12844012"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFGV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      395 a    290 c    284 g    380 t
ORIGIN
```

```
  Query Match             36.3%;  Score 460.2;  DB 12;  Length 1349;
  Best Local Similarity   63.7%;  Pred. No. 2.5e-123;
  Matches  699;  Conservative    0;  Mismatches  398;  Indels    0;  Gaps    0;

Qy     125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
           |||| ||||||  |||||||| ||||| | || ||     |    || |||||  ||||||
Db     132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy     185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
           |||||||||| |  |||||||||| || || ||    ||  | || ||||||   ||
Db     192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy     245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
           |  || |||| |  ||    || ||||||| |   |||||||||  |  |||
Db     252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy     305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |   || |  |||||||| | || ||||| ||||| ||||||| |||||||| ||||
Db     312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy     365 atgctggtttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
           ||||||| | ||| |||||| |||||||||| | || ||  | ||||| || ||| |
```

```
Db     372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy     425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                 ||   ||||  || ||| ||||||||||||||| | |||||| |||||||  |
Db     432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy     485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
           |||||||||  ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db     492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy     545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
           |||| ||  | || |||||||||||| || || |||||||| ||||| ||||  || |  |
Db     552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy     605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
           ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
Db     612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy     665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
           |||| |||   |  ||  |||   |  | ||  |  |  |  ||||||   | |||||
Db     672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784
           | |||   ||  ||   || | |   ||       |   |||| ||   |  | ||   |
Db     732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
           | || | | ||||   | || || ||     | ||  |  |  |||||||  |||  |
Db     792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy     845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
           | ||  | ||  |||| ||     |||| |||   |  || | ||  || |||||||||
Db     852 AAAACTTAAATGTGAGTCGCTTTGGTGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
           |||  |  || | ||||||   |||   |  || ||||  | || |  || | | |||
Db     912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy     965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
           |||| |||       | ||  ||  | | | |||||||  ||| |||      ||   |
Db     972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy     1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
            |   ||  ||||| || || |  ||| |||||||  || || || |||   ||  |  |||
Db     1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy     1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
            |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||| | ||||
Db     1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy     1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
              |   |  |||| |||| ||||||||||   |||||||  |||||     | |||||||
Db     1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy     1205 aaatcatccatctgatg 1221
           | ||  |     |||||
Db     1212 AGATAGTTACCATGATG 1228
```

```
RESULT   8
AK010139
LOCUS       AK010139     1350 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310069P19, full insert sequence.
ACCESSION   AK010139
VERSION     AK010139.1  GI:12845371
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310069P19.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 1350)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1350)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1350)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1350)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1350)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1350
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1898377"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310069P19"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53. .1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26725.1"
                     /db_xref="GI:12845372"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPIATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      398 a    289 c    283 g    380 t
ORIGIN
```

```
  Query Match           36.3%;  Score 460.2;  DB 12;  Length 1350;
  Best Local Similarity  63.7%;  Pred. No. 2.5e-123;
  Matches  699;  Conservative    0;  Mismatches  398;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||||  |||||||| ||||| | || ||     |    || |||||  ||||||
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          ||||||||||| |  |||||||||||| || || ||    ||  | || ||||||   ||
Db    192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db    252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||||| | || |||||  |||||  |||||||  ||||||||| ||||
Db    312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          |||||||  | ||| |||||| ||||||||||  | || ||  | |||||.|| ||| |
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    485 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 544
          |||||||||| ||    ||| |||| ||| | |||||||| || || || |||||   | |
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          |||| ||  | || |||||||||||| || || |||||||| |||||  |||| ||  | |
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||| | |||||||| | || || ||| ||||| || || || ||||| ||| || ||
Db    612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAATTGCTACTGTGAAGTATACAG 671

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          |||| |||   |  ||  |||   |  || |  |  |  |  ||||||   | |||| |
Db    672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||   ||  ||   || | |  ||       |   |||| ||   |  | | ||   |
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
```

```
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db   1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db   1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTTTGTACCATAAGGAGATTC 1151

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy   1205 aaatcatccatctgatg 1221
Db   1212 AGATAGTTACCATGATG 1228


RESULT  9
AK010124
LOCUS       AK010124     1340 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310069D09, full insert sequence.
ACCESSION   AK010124
VERSION     AK010124.1  GI:12845346
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310069D09.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1340)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1340)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1340)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1340)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1340)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1340
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1895811"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310069D09"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             47. .1234
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26715.1"
                     /db_xref="GI:12845347"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      393 a    287 c    282 g    378 t
ORIGIN


  Query Match           36.1%;  Score 458.6;  DB 12;  Length 1340;
  Best Local Similarity   63.6%;  Pred. No. 7.3e-123;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;
```

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||||  ||||||| |||||  | || ||      |    || |||||  ||||||
Db    126 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 185

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          ||||||||||  |  ||||||||||| || || ||     ||  | || ||||||   ||
Db    186 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 245

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
             |  || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db    246 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 305

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||||  | || ||||| ||||| |||||||  |||||||||  ||||
Db    306 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 365

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          |||||||  | |||  |||||| ||||||||||   |  || ||  | |||||  || |||  |
Db    366 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 425

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                ||   ||||  || ||| |||||||||||||  | ||||||  |||||||  |
Db    426 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 485

Qy    485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||||  ||    ||| |||| ||| | |||||||| || || || |||||   | |
Db    486 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 545

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcatttttccaccatgccagagc 604
          |||| ||  | || ||||||||||| || || |||||||| ||||| |||  || |  |
Db    546 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 605

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
Db    606 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 665

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          |||| |||   |  ||  |||  |  | || |  |  |  |||||   | ||||| |
Db    666 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 725

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||   ||  ||  || | |  ||      |  |||| ||   |  | ||   |
Db    726 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 785

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
          || | | || ||||   | || || ||     | ||  |  |  |||||||  |||  |
Db    786 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 845

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
          | ||  | ||  |||| ||       ||| |||   |  || | ||  || ||||||||||
Db    846 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 905

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
            ||| |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db    906 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 965

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| |||        | ||  ||  | | | |||||||  ||| |||       ||   |
Db    966 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1025

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          |  ||  ||||| || || |  ||| ||||||  || || || |||   ||  |  |||
Db   1026 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1085

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||||  | ||||
Db   1086 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1145

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
            |   |  |||| |||| ||||||||||   |||||||  |||||      | |||||||
Db   1146 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAGGAGGTTTACAATG 1205
```

```
Qy    1205 aaatcatccatctgatg 1221
           | ||  |     |||||
Db    1206 AGATAGTTACCATGATG 1222
```

RESULT 10
AK009479
LOCUS AK009479 1343 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310022K08, full insert sequence.
ACCESSION AK009479
VERSION AK009479.1 GI:12844303
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2310022K08.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1343)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 1343)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (bases 1 to 1343)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome research. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (bases 1 to 1343)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 1343)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5′ GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3′], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5′ GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3′]. cDNA was cleaved with XhoI and SstI. Cloning sites, 5′ end: XhoI; 3′ end: SstI. Host: SOLR.

FEATURES Location/Qualifiers
source 1. .1343
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1896992"
/db_xref="MGD:MGI:1914967"
/clone="2310022K08"
/sex="male"
/tissue_type="tongue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 50. .1237
/note="putative"
/codon_start=1
/protein_id="BAB26316.1"
/db_xref="GI:12844304"
/translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
AED"

BASE COUNT 394 a 289 c 282 g 378 t
ORIGIN

Query Match 36.1%; Score 458.6; DB 12; Length 1343;
Best Local Similarity 63.6%; Pred. No. 7.3e-123;
Matches 698; Conservative 0; Mismatches 399; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||||  |||||||  |||||  | || ||      |    || |||||  ||||||
Db    129 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 188

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          ||||||||||  |  ||||||||||| || || ||    ||  | || ||||||   ||
Db    189 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 248

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  || |||| |  ||    || |||||||  |  ||||||||||  |  |||
Db    249 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 308

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||||  | || |||||  |||||  |||||||  |||||||||  ||||
Db    309 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 368

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          |||||||  | |||  ||||||  ||||||||||.  | ||  ||  |  |||||  || |||  |
Db    369 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 428

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
              ||   ||||  || ||| ||||||||||||| | ||||||  |||||||  |
Db    429 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 488

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||||  ||    |||  ||||  |||  | ||||||||  ||  ||  ||  |||||  | |
```

```
Db    489 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 548

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          ||||  ||  |  ||  ||||||||||||  ||  ||  ||||||||  |||||  ||||  ||  |  |
Db    549 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 608

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          |||||  |  ||||||||  |  ||  ||  |||  |||||  ||   |  ||  |||||  |||  ||  ||
Db    609 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 668

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||||  |||    |  ||  |||    |  |  ||  |  |  |  ||||||    |  |||||  |
Db    669 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 728

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          |  |||   ||  ||   ||  |  |   ||       |   ||||  ||   |  |  ||    |
Db    729 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 788

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            ||  |  |  ||  ||||   |  ||  ||  ||     |  ||  |  |  |||||||  |||  |
Db    789 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 848

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
          |  ||  |  ||  ||||  ||       |||  |||    |  ||  |  ||  ||  ||||||||||
Db    849 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 908

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
            |||  |   ||  |  ||||||   |||    |  ||  |||||  |  ||  |  ||  |  |  |||
Db    909 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 968

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          ||||  |||      |  ||  ||  |  |  |  |||||||  |||  |||       ||     |
Db    969 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1028

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           |   ||  ||||| ||  ||  |  |||  ||||||  ||  ||  ||  |||    ||  |  |||
Db    1029 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1088

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           |  ||||  ||   |||||||||  |  |  |  |||||||  |  |||||||||  |  ||||
Db    1089 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1148

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           |   |   ||||  ||||  ||||||||||   |||||||  |||||    |  |||||||
Db    1149 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1208

Qy    1205 aaatcatccatctgatg 1221
           |  ||  |     |||||
Db    1209 AGATAGTTACCATGATG 1225
```

RESULT 11
AK009413
LOCUS AK009413 1345 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310020E22, full insert sequence.
ACCESSION AK009413
VERSION AK009413.1 GI:12844193
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2310020E22.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1345)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 1345)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

```
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1345)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1345)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1345)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES            Location/Qualifiers
     source         1. .1345
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /db_xref="MGD:MGI:1895774"
                    /db_xref="MGD:MGI:1914967"
                    /clone="2310020E22"
                    /sex="male"
                    /tissue_type="tongue"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="adult"
     CDS            52. .1239
                    /note="putative"
                    /codon_start=1
                    /protein_id="BAB26272.1"
                    /db_xref="GI:12844194"
                    /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                    SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                    FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                    GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                    FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                    YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                    AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                    AED"
BASE COUNT      395 a    288 c    283 g    379 t
ORIGIN
```

```
  Query Match             36.1%;  Score 458.6;  DB 12;  Length 1345;
  Best Local Similarity   63.6%;  Pred. No. 7.3e-123;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy     125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
           |||| ||||||  ||||||| ||||| | || ||      |    || |||||  ||||||
Db     131 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 190

Qy     185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
           ||||||||||| |  |||||||||||| || || ||    ||  | || ||||||   ||
Db     191 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 250

Qy     245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
               |  || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db     251 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 310

Qy     305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |  || |  |||||||| | || ||||| ||||| ||||||| |||||||| ||||
Db     311 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 370

Qy     365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
           ||||||| | ||| |||||| |||||||||| | || ||  | ||||| || ||| |
Db     371 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 430

Qy     425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
               ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db     431 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 490

Qy     485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
           |||||||||  ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db     491 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 550

Qy     545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
           |||| ||  | || ||||||||||| || || |||||||| ||||| ||||  || |  |
Db     551 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 610

Qy     605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
           ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
Db     611 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 670

Qy     665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
           |||| |||   |  ||  |||  |  | ||  |  |  | ||||||   | |||||  |
Db     671 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 730

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
           | |||   ||  ||   || | |   ||       |   |||| ||   |  | ||    |
Db     731 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 790

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
           || | | || ||||   | || || ||     | ||  |  |  ||||||  |||  |
Db     791 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 850

Qy     845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
           | ||  | ||  |||| ||     ||| |||   |  || | ||  || |||||||||
Db     851 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 910

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
```

```
          ||| |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db    911 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 970

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          ||||  |||       | ||  ||  | | | |||||||  |||  |||      ||    |
Db    971 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1030

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          |   ||  |||||  || || | . ||| ||||||  || || || |||   ||  |  |||
Db   1031 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1090

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  ||||  ||   |||||||||  |  |  |  |||||||  | |||||||||  | ||||
Db   1091 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1150

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           |   |    |||| |||| ||||||||||   |||||| |||||     | |||||||
Db   1151 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1210

Qy   1205 aaatcatccatctgatg 1221
          | ||  |     |||||
Db   1211 AGATAGTTACCATGATG 1227


RESULT  12
AK009473
LOCUS       AK009473     1345 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310022H17, full insert sequence.
ACCESSION   AK009473
VERSION     AK009473.1  GI:12844293
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310022H17.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1345)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1345)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1345)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1345)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1345)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1345
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1898391"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310022H17"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              52. .1239
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26312.1"
                      /db_xref="GI:12844294"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      395 a    288 c    283 g    379 t
ORIGIN


  Query Match           36.1%;  Score 458.6;  DB 12;  Length 1345;
  Best Local Similarity   63.6%;  Pred. No. 7.3e-123;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          ||||  ||||||  |||||||  |||||  |  ||  ||      |    ||  |||||  ||||||
Db    131 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 190

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          ||||||||||  |  ||||||||||| || || ||    ||  |  || ||||||   ||
Db    191 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 250
```

```
Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    251 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 310

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    311 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 370

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db    371 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 430

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    431 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 490

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
Db    491 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 550

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db    551 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 610

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
Db    611 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 670

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
Db    671 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 730

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
Db    731 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 790

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db    791 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 850

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    851 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 910

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db    911 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 970

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    971 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1030

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db   1031 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1090

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db   1091 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1150

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1151 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1210

Qy   1205 aaatcatccatctgatg 1221
Db   1211 AGATAGTTACCATGATG 1227
```

```
RESULT 13
AK009359
LOCUS       AK009359      1346 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310015H15, full insert sequence.
ACCESSION   AK009359
VERSION     AK009359.1  GI:12844108
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310015H15.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1346)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1346)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1346)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1346)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1346)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
```

through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

```
FEATURES             Location/Qualifiers
     source          1. .1346
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1895805"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310015H15"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53. .1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26240.1"
                     /db_xref="GI:12844109"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASAKNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKIHFIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      396 a    289 c    282 g    379 t
ORIGIN
```

Query Match 36.1%; Score 458.6; DB 12; Length 1346;
Best Local Similarity 63.6%; Pred. No. 7.3e-123;
Matches 698; Conservative 0; Mismatches 399; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||||  |||||||| ||||| | || ||     |    || |||||  ||||||
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          |||||||||| |  ||||||||||| || || ||    ||  | || ||||||   ||
Db    192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db    252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
            |   || |  |||||||||| | || ||||| ||||| ||||||| |||||||||| ||||
Db    312 CATCAGCCAAAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          ||||||| | ||| |||||| |||||||||   | || ||  | ||||| || ||| |
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
           |||||||| ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          |||| ||  | || ||||||||||| || || |||||||| ||||| ||||  || |  |
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
```

```
Db    612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          |||| |||   |  || ||   || | || |  |  |  |||||   | ||||| |
Db    672 AAAGTCCCTTTAAAAAGATTCACTTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784
          | |||   ||  ||   || | |   ||       |   |||| ||   |  | ||   |
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            || | | ||  ||||   | || || ||      | ||  |  |  |||||||  |||  |
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
          | ||  | ||  |||| ||      ||| |||   |  || | ||  || |||||||||
Db    852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
            ||| |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| |||       | || ||  | | | |||||||  ||| |||      ||   |
Db    972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           |   ||  ||||| || || |  ||| |||||  || || || |||   ||  |  |||
Db    1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||||  | ||||
Db    1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           |   |    |||| |||| ||||||||||   |||||| ||||    | |||||||
Db    1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy    1205 aaatcatccatctgatg 1221
           | ||  |     |||||
Db    1212 AGATAGTTACCATGATG 1228
```

```
RESULT  14
AK009523
LOCUS       AK009523     1346 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310028I09, full insert sequence.
ACCESSION   AK009523
VERSION     AK009523.1  GI:12844369
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310028I09.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1346)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1346)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1346)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
```

```
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1346)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1346)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES            Location/Qualifiers
     source         1. .1346
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /db_xref="MGD:MGI:1895759"
                    /db_xref="MGD:MGI:1914967"
                    /clone="2310028I09"
                    /sex="male"
                    /tissue_type="tongue"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="adult"
     CDS            53. .1240
                    /note="putative"
                    /codon_start=1
                    /protein_id="BAB26338.1"
                    /db_xref="GI:12844370"
                    /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                    SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                    FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                    GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                    FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                    YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                    AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                    AED"
BASE COUNT      395 a    289 c    283 g    379 t
ORIGIN
```

```
  Query Match           36.1%;  Score 458.6;  DB 12;  Length 1346;
  Best Local Similarity   63.6%;  Pred. No. 7.3e-123;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||||  |||||||  |||||  | ||  ||     |    ||  |||||  ||||||
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          ||||||||||  |  ||||||||||| || ||  ||     ||  | ||  ||||||   ||
Db    192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  ||  ||||  |  ||    ||  |||||||  |   ||||||||||  |  |||
Db    252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |  ||  |  ||||||||| |  ||  ||||  |||||  |||||||  |||||||| ||||
Db    312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          |||||||  | |||  ||||||  ||||||||||  | ||  ||  | |||||  || ||| |
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                ||   ||||  ||  |||  |||||||||||||||  | ||||||  |||||||  |
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||||  ||    |||  ||||  |||  |  ||||||||  ||  ||  ||  |||||  | |
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          ||||  |  | ||  ||||||||||||  ||  ||  ||||||||  |||||  ||||   ||  | |
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          |||||  |  ||||||||  |  ||  ||  |||  |||||  ||   |  ||  |||||  ||  ||  ||
Db    612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||||  |||   |  ||  |||   |  |  ||  |  |   |  |  ||||||   |  |||||  |
Db    672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||   ||  ||   ||  |  |   ||       |   ||||  ||   |  |  ||    |
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
          ||  |  |  ||  ||||   |  ||  ||  ||     |  ||   |   |  |||||||  |||  |
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
          |  ||   |  ||  ||||  ||       |||  |||    |   ||  |  ||  ||||||||||
Db    852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          |||  |   ||  |  ||||||   |||    |   ||  |||||   |  ||  |   ||  |  |||
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          ||||  |||        |  ||   ||  |  |  |  |||||||   |||  |||       ||    |
Db    972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
```

```
Db   1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  ||||  ||   |||||||||  |  |  |  |||||||  | |||||||||  | ||||
Db   1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           |   |    |||| |||| ||||||||||   |||||||  |||||     | |||||||
Db   1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy   1205 aaatcatccatctgatg 1221
          | ||  |     |||||
Db   1212 AGATAGTTACCATGATG 1228


RESULT 15
AK009729
LOCUS       AK009729     1346 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310040L03, full insert sequence.
ACCESSION   AK009729
VERSION     AK009729.1  GI:12844702
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310040L03.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1346)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1346)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1346)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1346)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1346)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers.
     source           1. .1346
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1898390"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310040L03"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              53. .1240
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26466.1"
                      /db_xref="GI:12844703"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      395 a    288 c    283 g    380 t
ORIGIN
```

```
  Query Match             36.1%;  Score 458.6;  DB 12;  Length 1346;
  Best Local Similarity   63.6%;  Pred. No. 7.3e-123;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy     125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
           |||| ||||||  |||||||  ||||| | || ||     |    || |||||  ||||||
Db     132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy     185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
           |||||||||| |  ||||||||||| || || ||    ||  | || ||||||   ||
Db     192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy     245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
           |  || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db     252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy     305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |   || |  |||||||| | || ||||| ||||| |||||| |||||||| |||| ||||
Db     312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371
```

```
Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          ||||||| | ||| |||||| |||||||||| | || || | ||||| || ||| |
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                ||   |||| || ||| |||||||||||||| | |||||| |||||||  |
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    485 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 544
           ||||||||| ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          |||| ||  | || |||||||||||| || || |||||||| ||||| |||| || | |
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
Db    612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          |||| |||   |  ||  ||| |  | ||  |  |  |  ||||||   | |||||
Db    672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||   ||  ||   || | |   ||       |   |||| ||   |  | ||    |
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
          || | | || ||||   | || || ||     | ||  |  |  |||||||  |||  |
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
          | ||  | ||  |||| ||      ||| |||   |  || | ||  || ||||||||||
Db    852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
            ||| |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| |||       | ||  ||  | | | |||||||  ||| |||      ||    |
Db    972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          |   ||  ||||| || || |  ||| ||||||  || || || |||   ||  |  |||
Db   1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||| | ||||
Db   1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
            |   |    |||| |||| |||||||||    |||||||  |||||     | |||||||
Db   1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy   1205 aaatcatccatctgatg 1221
          | ||  |     |||||
Db   1212 AGATAGTTACCATGATG 1228
```

Search completed: January 30, 2002, 21:37:55
Job time: 4437 sec